The Impact of Randomization in Smoothing Networks*

Marios Mavronicolas[†] *Thomas Sauerwald*[‡]

(October 11, 2008)

[*]A preliminary version of this work appeared in the *Proceedings of the 27th Annual ACM Symposium on Principles of Distributed Computing,* pp. 345–354, August 2008. This work has been partially supported by the IST Program of the European Union under contract number 15964 (AEOLUS) and by the German Science Foundation (DFG) Research Training Group GK-693 of the Paderborn Institute for Scientific Computation (PaSCo).

[†]Department of Computer Science, University of Cyprus, Nicosia CY-1678, Cyprus. Part of the work of this author was done while visiting the Faculty of Computer Science, Electrical Engineering and Mathematics, University of Paderborn. Email mavronic@cs.ucy.ac.cy

[‡]International Computer Science Institute, 1947 Center Street, Suite 600 Berkeley, CA 94704, USA. Part of the work of this author was done while at the Faculty of Computer Science, Electrical Engineering and Mathematics, University of Paderborn, and while visiting the Department of Computer Science, University of Cyprus. Email sauerwal@upb.de

Abstract

We revisit *randomized smoothing networks* [13], which are made up of *balancers* and *wires*. We assume that balancers are oriented independently and uniformly at random. *Tokens* arrive arbitrarily on w *input wires* and propagate asynchronously through the network; each token is served on the *output wire* it arrives at. The *smoothness* is the maximum discrepancy among the numbers of tokens arriving at the w output wires. We present a collection of lower and upper bounds on smoothness, which are to some extent surprising:

- The smoothness of a single *block network* is $\lg \lg w + \Theta(1)$ (with high probability), where the additive constant is between -2 and 3. This *tight* bound improves vastly over the upper bound of $\mathcal{O}(\sqrt{\lg w})$ from [13], and it essentially settles our understanding of the smoothing properties of the block network.
- Most significantly, the smoothness of the cascade of two block networks is no more than 17 (with high probability); this is the *first* known randomized network with so small depth ($2 \lg w$) and so good (constant) smoothness. The proof introduces some combinatorial and probabilistic structures and techniques which may be further applicable; the result demonstrates the full power of randomization in smoothing networks.
- There is no randomized 1-smoothing network of *width* w and *depth* d that achieves 1-smoothness with probability better than $\frac{\mathsf{d}}{w-1}$. In view of the *deterministic*, $\Theta(\lg w)$-depth, 1-smoothing network in [18], this result implies the *first* separation between deterministic and randomized smoothing networks.

These results demonstrate an unexpected limitation on the power of randomization for smoothing networks: although it yields constant smoothness using small depth, going down to smoothness of 1 requires *linear* depth.

Keywords: Smoothing metworks, balancing networks, randomization.

1 Introduction

1.1 Motivation and Framework

A *smoothing network* [5] is a distributed data structure which receives *tokens* issued arbitrarily by concurrent processes at w *input wires* and routes them asynchronously through a network to w *output wires*. The network consists of *balancers* and *wires*. A balancer is oriented either top or bottom; the first token through the balancer will be forwarded to its *top* or *bottom* output wire, respectively, and subsequent tokens will alternate. Each token represents a request by a *client* for a unit service provided by the *server* residing on the output wire the token arrives at. Tokens are dispersed through the network, thereby reducing *contention* (cf. [5, 10]). A *small-depth* smoothing network has depth polylogarithmic in w, thereby reducing *latency*.

The routing of tokens through the network must ensure that all servers receive approximately the same number of tokens, no matter how unbalanced the arrival of tokens on input wires is. The *smoothness* of the network is the maximum discrepancy among the numbers of tokens arriving at different output wires; a γ-*smoothing network* has smoothness γ. Small-depth smoothing networks with low smoothness are attractive for multiprocessor coordination and load balancing applications where low-contention and small-latency are required; these include *producers-consumers* [12, Section 1.3] and distributed numerical computations [7].

We only require *randomized initialization* [2, 13], where each balancer is oriented independently and uniformly at random in some initialization phase. The resulting networks were coined as *randomized smoothing networks* [2, 13]. Randomized initialization is distinguished from *arbitrary initialization* [14], where each balancer is oriented arbitrarily, while both are *local*. Local initialization withdraws the advantage of global consistency offered by *global initialization* [1, 5, 8, 16, 17, 18], where every balancer must be oriented in a certain way; however, it offers fault-tolerance against crashes, resets or replacements of balancers (cf. [22]).

Together with *counting networks* [5], smoothing networks have been studied quite extensively since their introduction in the seminal paper of Aspnes *et al.* [5]. Since early on, it has been a major challenge to construct small-depth smoothing networks with small smoothness (and counting networks) (cf. [8, 17, 18]). In this work, we report some progress on this challenge.

1.2 State-of-the-Art

Aspnes *et al.* [5, Theorems 3.6 and 4.4] presented the *first* counting (hence, 1-smoothing) networks with depth $\Theta(\lg^2 w)$; those were isomorphic to the *bitonic sorting network* due to Batcher [6] and the *periodic sorting network* due to Dowd *et al.* [9], respectively, with *comparators* replaced by balancers. These networks require global initialization.

Already in 1992, Klugerman and Plaxton [18] presented an elaborate random construction for an asymptotically efficient, $\Theta(\lg w)$-depth 1-smoothing network; the construction was later derandomized by Klugerman [17]. Unfortunately, their 1-smoothing network is impractical since it contains the AKS *(sorting) network* [3] due to Ajtai, Komlos and Szemeredi [3], which inherits huge constants*; moreover, their network requires global initialization.

Shortly after the work of Klugerman and Plaxton [17, 18], Aiello *et al.* [2] presented an elegant construction for a (partially) randomized 2-smoothing network, called *r-butterfly*, with depth $\lg w + o(\lg w)$ and no reliance on the AKS (sorting) network; however, global initialization is still required for some of the balancers in the r-butterfly.

More than a decade later, Herlihy and Tirthapura [13] considered the randomized *block network* [5, 9]. This is a very simple network of depth $\lg w$, which has been used in advanced constructions; for example, the periodic network is the cascade of $\lg w$ blocks [5, 9]. An upper bound of $2.36\sqrt{\lg w}$ (with high probability) on smoothness was shown in [13]; this bound is trivially inherited to the bitonic network [5, 6] and the periodic network [5, 9]. This upper bound improved over the smoothness of $\lg w$ known before for simple constructions with global initialization (such as the *bitonic merger* [16], the *butterfly* [17, 18]), and for the block network itself with (local) non-deterministic initialization [14].

We summarize all state-of-the-art results (together with the results presented in this article) on smoothing networks in Table 1.

1.3 Contribution and Significance

Herlihy and Tirthapura formulated in [13, Section 5] the following three interesting *Open Problems* about randomized smoothing networks, which we quote:

(1) Our bounds for the smoothness of the block network do not make use of structure that may be present in the input sequence. Can we obtain better bounds if the input is already fairly smooth?

(2) Can we get better bounds on the output smoothness of the randomized periodic or bitonic networks?

(3) How tight is the $\mathcal{O}(\sqrt{\lg w})$ upper bound for the block network? Can we get a matching lower bound?

*Knuth comments on the (im)practicality of the AKS sorting network with n inputs and n outputs as follows [19, p. 228]: "The networks they [Ajtai *et al.* [3]] constructed are not of practical interest, since many comparators were introduced just to save a factor of $\lg n$. Batcher's method is much better, unless n exceeds the total memory capacity of all computers on earth."

Network	Depth	Type	GI	Smoothness	With Probability	Reference
Bitonic	$\Theta(\lg^2 w)$	D	✓	1	Not applicable	[5, Theorem 3.6]
Periodic	$\Theta(\lg^2 w)$	D	✓	1	Not applicable	[5, Theorem 4.4]
KP	$\Theta(\lg w)$	D	✓	1	Not applicable	[18, Theorem 5.2]
r-Butterfly	$\lg w(1+o(1))$	D/R	✓	2	$\geq 1 - \frac{1}{\mathsf{superpoly}(w)}$	[2, Theorem 3.1]
Bitonic merger	$\lg w$	D	✓	$\lg w$	Not applicable	[16, Theorem 4.5]
Butterfly	$\lg w$	D	✓	$\lg w$	Not applicable	[18, Corollary 4.1.1]
Block	$\lg w$	D	X	$\lg w$	Not applicable	[14, Theorems 3 & 4]
Block	$\lg w$	R	X	$2.36\sqrt{\lg w}$	$\geq 1 - \frac{4}{w}$	[13, Theorem 10]
Block	$\lg w$	R	X	$\lceil \lg\lg w \rceil + 3$	$\geq 1 - \frac{4}{w^3}$	Theorem 5.1
Block	$\lg w$	R	X	$\lfloor \lg\lg w \rfloor - 2$	$\leq 2\exp\left(-\frac{4\sqrt{w}}{\lg w}\right)$	Theorem 5.3
Two Blocks	$2\lg w$	R	X	17	$\geq 1 - 2 \cdot \frac{4\lg\lg w - 39}{w}$	Theorem 6.7
Any	d	R	X	1	$\leq \frac{\mathsf{d}}{w-1}$	Theorem 7.1

Table 1: Summary of known bounds on the smoothness of smoothing networks. D and R stand for deterministic and randomized balancers, respectively; note that a deterministic balancer may result from either global or arbitrary initialization. D/R stands for a combination of deterministic and randomized balancers. GI stands for global initialization; the corresponding column indicates whether GI is required or not. KP stands for Klugerman and Plaxton [17, 18]. We use $\mathsf{superpoly}(w)$ to denote a superpolynomial function in w. Note that probabilities are not applicable in the case of deterministic balancers.

In this work, we shall provide answers to all three of these open problems.

1.3.1 One Block

We first prove that the randomized block network is $(\lceil \lg\lg w \rceil + 3)$-smoothing with probability at least $1 - \frac{4}{w^3}$ (Theorem 5.1). Our analysis drastically sharpens the elementary arguments developed by Herlihy and Tirthapura [13] for establishing their corresponding $\mathcal{O}(\sqrt{\lg w})$ upper bound. This exponential improvement is achieved through a tighter analysis of the same random variables. This result provides a partial answer to *Open Problem* (3) of Herlihy and Tirthapura from [13, Section 5].

For the proof, we consider a partition of the block into two groups of consecutive layers, and we analyze separately the contribution of each group to smoothness; the leftmost group consists of $\lg w - \lceil \lg\lg w \rceil$ layers. We consider a certain sum of independent random variables, which was bounded directly by using a *Hoeffding Bound* [15] in [13]; we now split this sum into two parts, which correspond to the two groups from the partition. Still using a Hoeffding

Bound, the first part is bounded by $\lg \lg w + \Theta(1)$; the second part is bounded by a constant by simply summing up the maximum possible absolute values of the involved random variables.

We continue to establish a matching lower bound (up to a small additive constant) on the smoothness of the block network. We prove that the randomized block network is only a $(\lfloor \lg \lg w \rfloor - 2)$-smoothing network with probability at most $2\,\mathsf{exp}\left(-\frac{4\sqrt{w}}{\lg w}\right)$ (Theorem 5.3); so, our analysis is essentially tight. This lower bound completes the answer to *Open Problem* (3) of Herlihy and Tirthapura from [13, Section 5].

For the proof of Theorem 5.3, we partition again the block network into two groups of consecutive layers. We determine a *fixed-point* input (cf. [14]) for the first group; we prove that (with probability at most $2\,\mathsf{exp}\left(-\frac{4\sqrt{w}}{\lg w}\right)$), this input achieves smoothness no better than $\lfloor \lg \lg w \rfloor - 1$ when traversing the second group.

Although there still remains a small gap between our upper and lower bounds on smoothness for a single block, we feel that these bounds essentially settle our understanding of the smoothing properties of the randomized block network with arbitrary inputs.

1.3.2 Two Blocks

As our main result, we provide a simple, $\Theta(\lg w)$-depth, randomized $\mathcal{O}(1)$-smoothing network. Specifically, we show that the cascade of *two* block networks is 17-smoothing with probability at least $1 - 2 \cdot \frac{4 \lg \lg w - 39}{w}$ (Theorem 6.7). This upper bound identifies the *first* $\Theta(\lg w)$-depth, randomized smoothing network which simultaneously *(i)* achieves constant smoothness, *(ii)* does *not* use the AKS network [3] (and, hence, it need not be inpractical), and *(iii)* does *not* require global initialization.

This upper bound provides an answer to *Open Problem* (1) of Herlihy and Tirthapura from [13]: When the input to a block network has the smoothing properties of a block's output, then the corresponding output is 17-smooth (with high probability).

For the proof, we consider a partition of the second cascaded block into no more than $\left\lceil \frac{1}{2} \lg \lg w \right\rceil - 6$ groups of consecutive layers (with an increasing number of layers per group). Each group is found to contribute an amplification of 1 to smoothness (while decreasing the corresponding probability). So, invoking Theorem 5.1 to the first cascaded block, implies a smoothness of 17 for the cascade. To establish the amplification, we employ some very delicate combinatorial and probabilistic structures and techniques; we believe that these will be instrumental elsewhere — for example, in showing that a small number of cascaded randomized blocks is 2-smoothing (with high probability), and we conjecture this to be the case:

Conjecture 1.1 *The cascade of a constant number of randomized block networks is a 2-smoothing network (with high probability).*

Since the cascade of two blocks is contained in the periodic network [5, 9], it follows that the randomized periodic network is a 17-smoothing network (with high probability) (Corollary 6.10). This settles *Open Problem* (2) of Herlihy and Tirthapura from [13].

1.3.3 Improbability of 1-Smoothing

We conclude with the *first* improbability result about randomized smoothing networks [2, 13]. We show that there is no randomized network of width w and depth d that achieves smoothness of 1 with probability greater than $\frac{\mathsf{d}}{w-1}$ (Theorem 7.1).

For the proof, we follow the *probabilistic method* [4]. We establish that on a certain random input, the output of a randomized network is 1-smooth with probability at most $\frac{\mathsf{d}}{w-1}$. This implies the existence of a *deterministic* input with this property, which settles the claim.

Theorem 7.1 is bad news: Any of the common (randomized) networks of depth $\mathcal{O}(\lg^2 w)$ (such as the bitonic [6, 5] and periodic networks [5, 9]) is 1-smoothing with extremely small probability. Even more so, only randomized smoothing networks of depth (no less than) *linear* in w may guarantee a smoothness of 1 with constant probability; hence, it is impossible to obtain a small-depth, randomized 1-smoothing network with constant probability (Corollary 7.4).

Recall the deterministic 1-smoothing network from [17, 18], which relies on the AKS network to achieve depth $\Theta(\lg w)$. So, Theorem 7.1 reveals the *first* separation between deterministic and randomized (1-)smoothing networks: There is a deterministic 1-smoothing network of depth $\Theta(\lg w)$, but no such randomized network (with high probability).

Theorems 6.7 and 7.1 demonstrate together a somehow unexpected limit on the power of randomization in smoothing networks: There is some threshold c between 1 and 16 such that there are small-depth, randomized $(c+1)$-smoothing networks (with high probability), but no such c-smoothing networks. (Conjecture 1.1 postulates that $c = 1$.)

1.4 Road Map

The rest of this paper is organized as follows. Section 2 collects together some preliminary facts and notation. Smoothing networks are reviewed in Section 3. Section 4 revisits the block network (and some relatives). Sections 5 and 6 treat one block and the cascade of two blocks, respectively. The improbability result is presented in Section 7. We conclude, in Section 8, with a discussion of our results and some open problems.

2 Notation and Preliminaries

2.1 Notation

All logarithms are to the base 2. For an integer $i \geq 0$, the binary representation of i is a binary string $i_1 i_2 \ldots i_r$ with $r \geq \lceil \lg i \rceil$ such that $\sum_{k=1}^{r} 2^{r-k} i_k = i$. We shall often identify the integer i with the binary representation $i_1 i_2 \ldots i_r$ and write $i = i_1 i_2 \ldots i_r$; for each index ℓ with $1 \leq \ell \leq r$, $i(\ell)$ denotes the binary string $i_1 \ldots i_{\ell-1} \overline{i_\ell}, i_{\ell+1} i_r$ where bit ℓ has been reversed. For a binary string i, denote as $1(i)$ the number of occurrences of 1 in i. For an integer $i \geq 1$, denote $[i] = \{0, \ldots, i-1\}$. For a number $r \in \mathbb{R}$, denote $\mathsf{exp}(r) = e^r$.

An integer vector $\langle x_0, \ldots, x_{w-1} \rangle$ with $w \geq 2$ will be denoted as $\mathbf{x}$; the index of each entry will be represented as a binary string. For a vector $\mathbf{x}$, denote $\sum \mathbf{x} = \sum_{i \in [w]} x_i$; for an index set $\mathcal{I} \subseteq [w]$, denote $\sum_{\mathcal{I}} \mathbf{x} = \sum_{i \in [\mathcal{I}]} x_i$. The *odd-characteristic* function of the vector $\mathbf{x}$, denoted as $\mathsf{Odd}(\mathbf{x})$, is given by $\mathsf{Odd}(\mathbf{x}) = 1$ if $\sum \mathbf{x} = \sum_{i \in [w]} x_i$ is odd, and 0 otherwise. For a vector $\mathbf{x}$ with $w \geq 4$ entries, the ***A-cochain*** (resp., ***B-cochain***) of $\mathbf{x}$, denoted as $\mathbf{x}_A$ (resp., $\mathbf{x}_B$) consists of all entries of $\mathbf{x}$ whose indices have least significant bits 00 or 11 (resp., 01 or 10).

We shall use $\mathbb{P}[\mathcal{E}]$ to denote the *probability* of the *event* $\mathcal{E}$ in some probability space. We shall denote the complement of event $\mathcal{E}$ as $\neg\mathcal{E}$. Say that the events $\mathcal{E}_1$ and $\mathcal{E}_2$ are *conditionally independent given the event* $\mathcal{E}$ if $\mathbb{P}[\mathcal{E}_1 \wedge \mathcal{E}_2 \mid \mathcal{E}] = \mathbb{P}[\mathcal{E}_1 \mid \mathcal{E}] \cdot \mathbb{P}[\mathcal{E}_2 \mid \mathcal{E}]$ (cf. [11, Section 21.5]). All random variables considered in this paper will be discrete and have a finite range. For a random variable v, we shall denote as $\mathbb{E}[v]$ the *expectation* of v (according to $\mathbb{P}$); $\mathsf{Range}(v)$ will denote the *range* of v and $|\mathsf{Range}(v)|$ will denote its size.

2.2 Smooth and Concentrated Vectors

Fix a number $\gamma \geq 1$. Say that the vector $\mathbf{x}$ is ***γ-smooth*** if for each pair of distinct entries x_i and x_j with $i, j \in [w]$, $|x_i - x_j| \leq \gamma$; so, all entries are within γ of each other. Say that $\mathbf{x}$ is ***ceiling γ-concentrated*** if for each entry x_i with $i \in [w]$, $x_i \leq \left\lceil \frac{\sum \mathbf{x}}{w} \right\rceil + \gamma$; so, the entry x_i exceeds the ***ceiling average*** $\left\lceil \frac{\sum \mathbf{x}}{w} \right\rceil$ by no more than γ. Say that $\mathbf{x}$ is ***floor γ-concentrated*** if for each entry x_i with $i \in [w]$, $x_i \geq \left\lfloor \frac{\sum \mathbf{x}}{w} \right\rfloor - \gamma$; so, the entry x_i is exceeded by the ***floor average*** $\left\lfloor \frac{\sum \mathbf{x}}{w} \right\rfloor$ by no more than γ. Say that $\mathbf{x}$ is ***ceilingfloor γ-concentrated*** if it is both ceiling γ-concentrated and floor γ-concentrated. Finally, say that $\mathbf{x}$ is ***γ-concentrated*** if for each entry x_i with $i \in [w]$, $\left| x_i - \frac{\sum \mathbf{x}}{w} \right| < \gamma$; so, the entry x_i is within less than γ of the ***average*** $\frac{\sum \mathbf{x}}{w}$.

In the next two simple claims, we shall observe that (ceilingfloor) concentration implies smoothness. We first prove:

Lemma 2.1 *Assume that* $\mathbf{x}$ *is ceilingfloor* γ*-concentrated. Then,* $\mathbf{x}$ *is* $(2\gamma+1)$*-smooth.*

Proof: Fix an arbitrary pair of entries x_j and x_k with $j, k \in [w]$. Since $\mathbf{x}$ is ceilingfloor γ-concentrated, $\left\lfloor \frac{\sum \mathbf{x}}{w} \right\rfloor - \gamma \leq x_j \leq \left\lceil \frac{\sum \mathbf{x}}{w} \right\rceil + \gamma$ and $\left\lfloor \frac{\sum \mathbf{x}}{w} \right\rfloor - \gamma \leq x_k \leq \left\lceil \frac{\sum \mathbf{x}}{w} \right\rceil + \gamma$. It follows that

$$\begin{aligned} |x_j - x_k| &= \left| \left(x_j - \left\lceil \frac{\sum \mathbf{x}}{w} \right\rceil \right) + \left(\left\lfloor \frac{\sum \mathbf{x}}{w} \right\rfloor - x_k \right) + \left(\left\lceil \frac{\sum \mathbf{x}}{w} \right\rceil - \left\lfloor \frac{\sum \mathbf{x}}{w} \right\rfloor \right) \right| \\ &\leq |2\gamma + 1| \\ &= 2\gamma + 1 \,, \end{aligned}$$

as needed. ■

We continue to prove:

Lemma 2.2 *Assume that* $\mathbf{x}$ *is* γ*-concentrated, where* 2γ *is an integer.. Then,* $\mathbf{x}$ *is* $(2\gamma-1)$*-smooth.*

Proof: Fix an arbitrary pair of entries x_j and x_k with $j, k \in [w]$. Since $\mathbf{x}$ is γ-concentrated, $\frac{\sum \mathbf{x}}{w} - \gamma < x_j < \frac{\sum \mathbf{x}}{w} + \gamma$ and $\frac{\sum \mathbf{x}}{w} - \gamma < x_k < \frac{\sum \mathbf{x}}{w} + \gamma$. By the triangle inequality,

$$\begin{aligned} |x_j - x_k| &= \left| \left(x_j - \frac{\sum \mathbf{x}}{w} \right) + \left(\frac{\sum \mathbf{x}}{w} - x_k \right) \right| \\ &\leq \left| x_j - \frac{\sum \mathbf{x}}{w} \right| + \left| \frac{\sum \mathbf{x}}{w} - x_k \right| \\ &< 2\gamma \,. \end{aligned}$$

Since 2γ is an integer, it follows that

$$|x_j - x_k| \leq 2\gamma - 1 \,,$$

as needed. ■

2.3 Probabilistic Tools and Preliminaries

In some later proofs, we shall use extensively the elementary ***Union Bound*** and the ***Hoeffding Bound*** [15]:

Lemma 2.3 (Union Bound) *For a finite sequence of events $\mathcal{E}_1, \mathcal{E}_2, \ldots$,*

$$\mathbb{P}\left[\bigvee_{i\geq 1} \mathcal{E}_i\right] \leq \sum_{i\geq 1} \mathbb{P}\left[\mathcal{E}_i\right] .$$

Lemma 2.4 (Hoeffding Bound) *Consider a collection of independent random variables $v_i \in [a_i, b_i]$ with $i \in [n]$. Then, for any number $\delta \geq 0$,*

$$\mathbb{P}\left[\left|\sum_{i=1}^{n} v_i - \mathbb{E}\left[\sum_{i=1}^{n} v_i\right]\right| \geq \delta\right] \leq 2 \cdot \mathsf{exp}\left(-\frac{2\delta^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right) .$$

The next observation relates the probability of the smoothing property for a random vector to the ones of the ceiling and floor concentration properties for the same vector.

Lemma 2.5 *Consider a random vector $\mathbf{x}$ such that $\mathbb{P}\left[\mathbf{x}\text{ is ceiling }\gamma\text{-concentrated}\right] \geq 1 - \delta$ and $\mathbb{P}\left[\mathbf{x}\text{ is floor }\gamma\text{-concentrated}\right] \geq 1 - \delta$ Then,*

$$\mathbb{P}\left[\mathbf{x}\text{ is }(2\,\gamma + 1)\text{-smooth}\right] \geq 1 - 2\,\delta ,$$

Proof: By Lemma 2.1 and the Union Bound (Lemma 2.3),

$$\begin{aligned}
\mathbb{P}\left[\mathbf{x}\text{ is }(2\gamma+1)\text{-smooth}\right] &\geq \mathbb{P}\left[\mathbf{x}\text{ is ceilingfloor }\gamma\text{-concentrated}\right] \\
&= 1 - \mathbb{P}\left[\mathbf{x}\text{ is not ceilingfloor }\gamma\text{-concentrated}\right] \\
&= 1 - \mathbb{P}\left[\mathbf{x}\text{ is not ceiling }\gamma\text{-concentrated or }\mathbf{x}\text{ is not floor }\gamma\text{-concentrated}\right] \\
&\geq 1 - \mathbb{P}\left[\mathbf{x}\text{ is not ceiling }\gamma\text{-concentrated}\right] - \mathbb{P}\left[\mathbf{x}\text{ is not floor }\gamma\text{-concentrated}\right] \\
&= -1 + \mathbb{P}\left[\mathbf{x}\text{ is ceiling }\gamma\text{-concentrated}\right] + \mathbb{P}\left[\mathbf{x}\text{ is floor }\gamma\text{-concentrated}\right] \\
&\geq -1 + (1-\delta) + (1-\delta) \\
&= 1 - 2\,\delta ,
\end{aligned}$$

as needed. ■

We conclude with an elementary technical claim, which will be used in a later proof.

Claim 2.6 *For a pair of random variables v_1 and v_2, and a pair of numbers δ_1 and δ_2,*

$$\mathbb{P}\left[v_1 + v_2 \geq \delta_1 + \delta_2\right] \leq \mathbb{P}\left[v_1 \geq \delta_1\right] + \mathbb{P}\left[v_2 > \delta_2\right] .$$

Proof: Clearly, the event $(v_1 < \delta_1) \wedge (v_2 \leq \delta_2)$ implies the event $(v_1 + v_2 < \delta_1 + \delta_2)$. By contrapositive, the event $\neg(v_1 + v_2 < \delta_1 + \delta_2) = (v_1 + v_2 \geq \delta_1 + \delta_2)$ implies the event $\neg\left((v_1 < \delta_1) \wedge (v_2 \leq \delta_2)\right) = \neg(v_1 < \delta_1) \vee \neg(v_2 \leq \delta_2) = (v_1 \geq \delta_1) \vee (v_2 > \delta_2)$. Hence, by the Union Bound (Lemma 2.3),

$$\begin{aligned}
\mathbb{P}\left[v_1 + v_2 \geq \delta_1 + \delta_2\right] &\leq \mathbb{P}\left[(v_1 \geq \delta_1) \vee (v_2 > \delta_2)\right] \\
&\leq \mathbb{P}\left[v_1 \geq \delta_1\right] + \mathbb{P}\left[v_2 > \delta_2\right] ,
\end{aligned}$$

as needed. ■

3 Smoothing Networks

A smoothing network [5] is a special case of a balancing network [5], which is a collection of interconnected balancers. Balancers and balancing networks are considered in Sections 3.1 and 3.2, respectively. Sections 3.3 and 3.4 consider deterministic and randomized balancing networks, respectively. Most definitions are given in the style of [13, 14].

3.1 Balancers

A ***balancer*** [5] is an asynchronous switch b with two ***input wires*** and two ***output wires***, denoted as $i_1(\mathsf{b})$ and $i_2(\mathsf{b})$ with $i_1(\mathsf{b}) < i_2(\mathsf{b})$; we shall often identify b with the set $\{i_1(\mathsf{b}), i_2(\mathsf{b})\}$. An ***initialization*** takes places in some preprocessing phase; the initialization simply chooses an ***orientation*** for the balancer b, either top or bottom. If the balancer is oriented top (resp. bottom), then the output wire $i_1(\mathsf{b})$ is labeled top (resp., top), the output wire $i_2(\mathsf{b})$ is labeled bottom (resp., top), and we write $i_2(\mathsf{b}) \stackrel{\mathsf{b}}{\rightarrow} i_1(\mathsf{b})$ (resp., $i_1(\mathsf{b}) \stackrel{\mathsf{b}}{\rightarrow} i_2(\mathsf{b})$).

A stream of tokens enters a balancer via its two input wires; each time a new token enters on an input wire, it is directed to the output wire currently labeled top; at the same time, the orientation of the balancer is reversed (from top to bottom and vice versa). This ensures that the total number of tokens is (almost) evenly balanced among the two output wires. A balancer satisfies a ***safety*** and a ***liveness condition*** [5]: it never creates or "swallows" tokens, respectively. Henceforth, we shall focus on *finite* streams of entering tokens; then, the liveness property implies that the balancer reaches a ***quiescent state*** [5] where all entering tokens have exited. Henceforth, we shall only consider a balancer in a quiescent state.

For a balancer b (in a quiescent state), denote as x_1 and x_2 the number of tokens entering on the input wires $i_1(\mathsf{b})$ and $i_2(\mathsf{b})$, of b, respectively; x_1 and x_2 make the ***input vector*** $\mathbf{x}_{\mathsf{b}} = \langle x_1, x_2 \rangle$. Denote as y_1 and y_2 the number of tokens exiting through the output wires $i_1(\mathsf{b})$ and $i_2(\mathsf{b})$ of b, respectively; y_1 and y_2 make the ***output vector*** $\mathbf{y}_{\mathsf{b}} = \langle y_1, y_2 \rangle$. We shall sometimes write $x_1(\mathsf{b})$, $x_2(\mathsf{b})$, $y_1(\mathsf{b})$ and $y_2(\mathsf{b})$ for x_1, x_2, y_1 and y_2, respectively, when reference to b is necessary; write $x(\mathsf{b})$ to denote either $x_1(\mathsf{b})$ or $x_2(\mathsf{b})$. If b is oriented top (resp., bottom), then $y_1 = \left\lceil \frac{x_1 + x_2}{2} \right\rceil$ and $y_2 = \left\lfloor \frac{x_1 + x_2}{2} \right\rfloor$ (resp., $y_1 = \left\lfloor \frac{x_1 + x_2}{2} \right\rfloor$ and $y_2 = \left\lceil \frac{x_1 + x_2}{2} \right\rceil$). Note that in all cases, $|y_2 - y_1| \leq 1$; so, the output vector of a balancer is 1-smooth.

3.2 Balancing Networks

A ***balancing network*** [5], or ***network*** for short, is an acyclic network of balancers, where output wires of balancers are connected to input wires of (other) balancers. The ***input wires*** $0, 1, \ldots, w-1$ may not be connected from any output wires; the ***output wires*** may not be

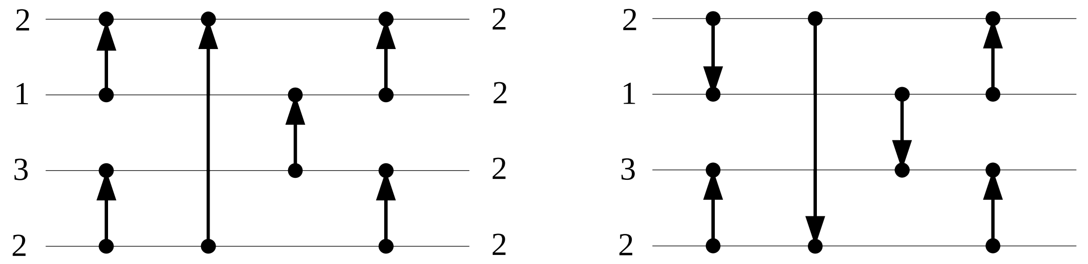

Figure 1: Two different orientations of the same balancing network. Horizontal segments represent the wires and vertical segments represent the balancers. Input and output wires appear on the left and right, respectively. The numbers next to the input and output wires denote the numbers of tokens on the input and output wires, respectively.

connected to any input wires. We shall consider a balancing network B_w with the same number w of input and output wires, called the network's ***width***. The ***balancer set*** of the network B_w, denoted as $\mathcal{B}_w$, is the set of all balancers in B_w. The ***size*** of the balancing network B_w, denoted as $\mathsf{size}(\mathsf{B}_w)$, is the number of its balancers; so, $\mathsf{size}(\mathsf{B}_w) = |\mathcal{B}_w|$. For an integer $\kappa \geq 1$, B_w^κ denotes the sequential cascade of κ copies of the network B_w.

3.2.1 Initializations and Orientations

We make a distinction on the way the balancers are initialized; in all cases, an initialization determines an orientation for each balancer in the network. In ***global initialization*** [5], each balancer must be oriented in some certain way. In ***local initialization*** [2, 13, 14], each balancer is oriented *locally* and *independently* of other balancers (cf. Section 3.4).

An ***orientation*** or for a network is a collection of orientations, one for each of its balancers; see Figure 1 for an illustration. There are $2^{\mathsf{size}(\mathsf{B}_w)}$ orientations for the network B_w. In the ***standard orientation***, all balancers are oriented top.

3.2.2 Layers and Paths

The acyclicity ensures that each balancer can be assigned a unique ***layer***: the length of the longest path from an input wire to the balancer; the ***depth*** $\mathsf{d}(\mathsf{B}_w)$ is the maximum layer. We shall often identify a layer with a corresponding set of balancers (and sometimes with the set of output wires of balancers in the layer). For a wire $i \in [w]$ and a layer ℓ with $1 \leq \ell \leq \mathsf{d}(\mathsf{B}_w)$, $\mathsf{b}_i(\ell)$ denotes the balancer b in layer ℓ such that $i \in \mathsf{b}$ (or $\perp$ if there is no such balancer). For an even integer $w \geq 2$, a layer ℓ is ***complete*** if it has $\frac{w}{2}$ balancers; the network B_w is ***complete*** if all layers are complete. An example of a complete layer is the ***ladder*** layer L_w, where there is a balancer connecting each wire $i \in \{0,1\}^{\lg w}$ to wire $i(1)$; so, $\mathsf{d}(\mathsf{L}_w) = 1$.

Fix an integer ℓ with $1 \leq \ell \leq \mathsf{d}(\mathsf{B}_w)$. The ***prefix network*** $\mathsf{Pref}_\ell(\mathsf{B}_w)$ consists of the ℓ leftmost layers $1, \ldots, \ell$ of the network B_w; the ***suffix network*** $\mathsf{Suff}_\ell(\mathsf{B}_w)$ consists of its ℓ rightmost layers $\mathsf{d}(\mathsf{B}_w) - \ell + 1, \ldots, \mathsf{d}(\mathsf{B}_w)$. The ***shift network*** $\mathsf{Shift}_\ell(\mathsf{B}_w)$ consists of layers $\ell, \ell+1, \ldots, \mathsf{d}(\mathsf{B}_w), 1, \ldots, \ell - 1$; so, $\mathsf{Shift}_\ell(\mathsf{B}_w)$ is a cyclic shift of the layers of the network B_w. More generally, for a pair of integers ℓ_1 and ℓ_2 with $\ell_1 + \ell_2 < \mathsf{d}(\mathsf{B}_w)$, the ***restriction*** $\mathsf{B}_w \setminus [\ell_1, \ell_2]$ of the network B_w consists of layers $\ell_1 + 1, \ldots, \mathsf{d}(\mathsf{B}_w) - \ell_2$; so, $\mathsf{B}_w \setminus [\ell_1, \ell_2] = \mathsf{Pref}_{\mathsf{d}(\mathsf{B}_w) - \ell_1 - \ell_2}\left(\mathsf{Suff}_{\mathsf{d}(\mathsf{B}_w) - \ell_1}(\mathsf{B}_w)\right)$.

A ***path*** (from b_1 to b_κ) is a sequence $\beta = \mathsf{b}_1, \ldots, \mathsf{b}_\kappa$ of interconnected balancers; so, an output wire of each balancer (other than the last) in the sequence is an input wire to the following balancer. Associated with β is a ***wire path***, which is the sequence of all output wires of balancers in the path β, where each such wire either connects to the following balancer in β or is an output wire of the network. For each balancer b_r with $1 < r \leq \kappa$, $x(\mathsf{b}_r)$ denotes the input to balancer b_r from balancer b_{r-1}; while $\widehat{x}(\mathsf{b}_r)$, denotes the other input to b_r; for $r = 1$, $x(\mathsf{b}_1)$ is arbitrarily chosen from $x_1(\mathsf{b}_1)$ and $x_2(\mathsf{b}_1)$.

A ***path descriptor*** for the wire path associated with β is a binary sequence of length κ such that a 0 (resp. 1) in position r indicates that balancer b_{r+1} (or an output wire of the network if $r = \kappa$) is reached through the top (resp., bottom) output wire of the preceding balancer b_r.

For a pair of layers ℓ and $\ell' < \ell$, $\mathsf{P}[\ell', \ell]$ denotes the set of all paths from layer ℓ' to layer ℓ. By abuse of notation, $\mathsf{P}[\mathsf{B}_w]$ denotes the set of all paths from an input wire to an output wire of the network B_w; so, $\mathsf{P}[\mathsf{B}_w] = \mathsf{P}[1, \mathsf{d}(\mathsf{B}_w)]$. Note that all paths from an input wire to an output wire of a complete balancing network B_w have length $\mathsf{d}(\mathsf{B}_w)$.

3.2.3 Dependencies

A balancer b in layer ℓ ***depends*** on balancer b' in layer $\ell' \leq \ell$ if there is a path from b' to b; by convention, the balancer b depends trivially on itself. Note that the balancer b may not depend on any other balancer in layer ℓ. We shall often abuse notation to say that each output wire of balancer b ***depends*** on balancer b' whenever b depends on b'. Consider two output wires j_1 and j_2 in layer ℓ; say that j_1 and j_2 are ***independent for layer*** $\ell' < \ell$ if there is no balancer b' in layer ℓ' such that both j_1 and j_2 depend on b'.

The ***dependency set*** of balancer b in layer ℓ is the set of all balancers b' in layers $\ell' \leq \ell$ such that b depends on b'. The ***input dependency set*** of balancer b in layer ℓ is the set of input wires of balancers in the dependency set of balancer b. We shall often abuse notation to identify the dependency set and the input dependency set of an output wire of the balancer b with the dependency set and the input dependency set, respectively, of b.

Fix a pair of layers ℓ and $\ell' \leq \ell$. The ***dependency set in*** ℓ' of an output wire j (of a balancer b) in layer ℓ is the restriction of the dependency set of j to balancers in layer ℓ'; denote it as $\mathsf{D}_j[\ell']$. The ***dependency set back to*** ℓ' of an output wire j (of a balancer b) in layer ℓ is the restriction of the dependency set of j to balancers in layers $\ell', \ldots, \ell$; denote it as $\mathsf{D}_j[\ell', \ell]$. The ***input dependency set in*** ℓ' of the output wire j (of a balancer b) in layer ℓ is the restriction of the input dependency set of j to input wires in layer ℓ'; denote it as $\mathsf{ID}_j[\ell']$. The ***input dependency set back to*** ℓ' of the output wire j of a balancer b in layer ℓ is the restriction of the input dependency set of j to input wires in layers $\ell', \ldots, \ell$; denote it as $\mathsf{ID}_j[\ell', \ell]$.

The balancing network B_w is ***full*** if for each layer ℓ with $1 \leq \ell \leq \mathsf{d}(\mathsf{B}_w)$, for each output wire j (of balancer b) in layer ℓ, for each layer $\ell' \leq \ell$, $|\mathsf{D}_j[\ell', \ell]| = 2^{\ell-\ell'}$; in such case, $|\mathsf{ID}_j[\ell', \ell]| = 2^{\ell-\ell'+1}$. A simple necessary condition for a full balancing network is that a pair of balancers b and b' in layers ℓ and $\ell' < \ell$, respectively, are either connected by a unique path or not connected at all. Note that each shift of a full balancing network is also full.

3.2.4 Topological Equivalence

Two balancing networks B_w and $\widehat{\mathsf{B}}_w$ with $\mathsf{d}(\mathsf{B}_w) = \mathsf{d}(\widehat{\mathsf{B}}_w)$ are ***topologically equivalent*** if there is a permutation $\pi : [w] \rightarrow [w]$ such that for each layer ℓ with $1 \leq \ell \leq \mathsf{d}(\mathsf{B}_w)$, there is a balancer b between wires j and k in B_w if and only if there is a balancer between wires $\pi(j)$ and $\pi(k)$ in $\widehat{\mathsf{B}}_w$; note that the permutation π induces a corresponding permutation ρ between the sets of balancers of the two networks B_w and $\widehat{\mathsf{B}}_w$. Roughly speaking, two balancing networks are topologically equivalent if there is a permutation between the wires of the two networks that yields one network from the other; call it a ***balancer-preserving*** permutation.

Note that topological equivalence is defined with no regard to the orientations of corresponding balancers in the two networks. Two balancing networks are ***isomorphic*** if they are topologically equivalent with a permutation ρ such that a balancer b in the network B_w is oriented top if and only if the balancer $\rho(\mathsf{b})$ in the network $\widehat{\mathsf{B}}_w$ is oriented top. Roughly speaking, two networks are isomorphic if the induced permutation between their sets of balancers respects the orientation of each balancer; so, isomorphism is a more stringent property than topological equivalence.

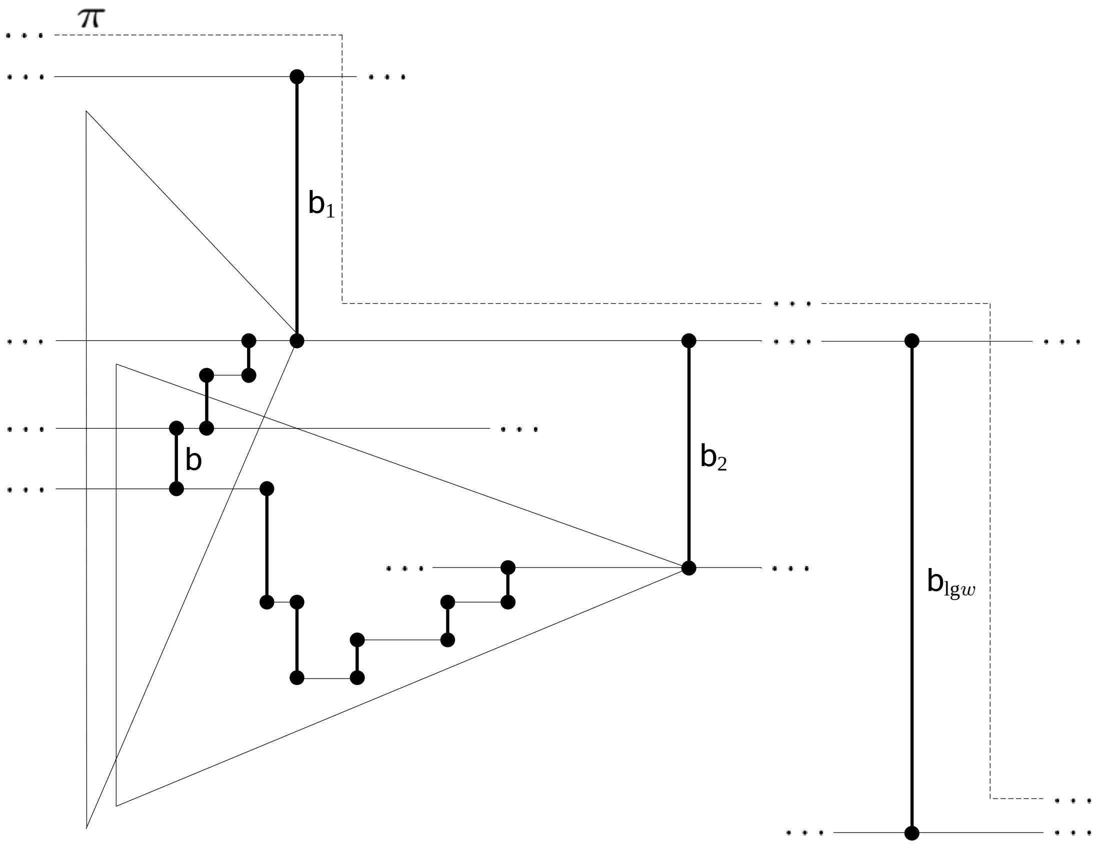

Figure 2: The path β and the balancers b, b_1 and b_2 involved in the proof of Lemma 3.3.

3.2.5 Delta and Bidelta Networks

A ***delta network*** [20] is one in which *(i)* there is a unique path from each input wire to every output wire, and *(ii)* the path descriptors associated with all paths ending with the same output wire are identical. A ***bidelta network*** [20] is one that is a delta network in both directions (from left to right and vice versa). It is known that any two bidelta networks of the same width and depth are topologically equivalent [20]. We continue with a simple observation.

Claim 3.1 *Consider a bidelta network* B*, and fix a pair of balancers* b_1 *and* b_2 *with* $\ell(\mathsf{b}_1) < \ell(\mathsf{b}_2)$ *on a path* β*. Then, the dependency sets* $\mathsf{D}_{\hat{x}(\mathsf{b}_1)}[1, \ell(\mathsf{b}_1)]$ *and* $\mathsf{D}_{\hat{x}(\mathsf{b}_2)}[1, \ell(\mathsf{b}_2)]$ *are disjoint.*

Figure 2 provides an illustration to the proof.

Proof: Assume, by way of contradiction, that the sets $\mathsf{D}_{\hat{x}(\mathsf{b}_1)}[1, \ell(\mathsf{b}_1)]$ and $\mathsf{D}_{\hat{x}(\mathsf{b}_2)}[1, \ell(\mathsf{b}_2)]$ are not disjoint. Then, there is a balancer $\mathsf{b} \in \mathsf{D}_{\hat{x}(\mathsf{b}_1)}[1, \ell(\mathsf{b}_1)] \cap \mathsf{D}_{\hat{x}(\mathsf{b}_2)}[1, \ell(\mathsf{b}_2)]$; take such a balancer

b with a maximum layer. By definition of dependency set, there is a path β_1 from b to b_2 and a path β_2 from b to b_1 in B. Since $\ell(\mathsf{b})$ is maximum, it follows that the paths β_1 and β_2 do not intersect. Consider now the concatenation $\beta_1\beta[\ell_1, \ell_2]$; clearly, it yields an alternative (to β_2) path from b_1 to b_2. Since B is a bidelta network, there is a unique path from b_1 to b_2. A contradiction. ■

3.3 Deterministic Balancing Networks

A ***deterministic balancer*** is one that has been oriented in some *fixed* way, as a result of either global initialization [5] or local initialization [2, 13, 14] (cf. Section 3.2.1). A ***deterministic balancing network***, or ***deterministic network*** for short, consists of deterministic balancers.

Consider an ***input vector*** $\mathbf{x} = \langle x_0, \ldots, x_{w-1}\rangle$, where x_i is the number of tokens fed into input wire i of the network B_w. The safety and liveness conditions for a balancer imply corresponding safety and liveness conditions for the network B_w in the natural way; in particular, the network B_w reaches a ***quiescent state*** on the input vector $\mathbf{x}$, where all $\sum \mathbf{x}$ input tokens have exited. It is simple to observe that on the input vector $\mathbf{x}$, the network B_w will reach a quiescent state with a (unique) ***output vector*** $\mathbf{y} = \langle y_0, \ldots, y_{w-1}\rangle$ (cf. [18, Lemma 3.1]). So, the network B_w acts like an operator $\mathsf{B} : \mathbf{x} \rightarrow \mathbf{y}$ transforming an input vector $\mathbf{x}$ into the output vector $\mathbf{y} = \mathsf{B}_w(\mathbf{x})$. A vector $\mathbf{x}$ is a ***fixed-point*** for the network B_w if $\mathsf{B}_w(\mathbf{x}) = \mathbf{x}$ (cf. [14]).

For each layer ℓ with $1 \leq \ell \leq \mathsf{d}(\mathsf{B}_w)$ in the network B_w, denote as $\mathbf{x}(\ell)$ and $\mathbf{y}(\ell)$ the input and output vectors, respectively, for layer ℓ; so, $\mathbf{x}(1) = \mathbf{x}$ and $\mathbf{y}(\mathsf{d}(\mathsf{B}_w)) = \mathbf{y}$. Clearly, $\sum \mathbf{y}(\ell) = \sum \mathbf{x}(\ell)$ for each layer ℓ with $1 \leq \ell \leq \mathsf{d}(\mathsf{B}_w)$.

3.3.1 Smoothing Networks and Path-Concentrating Networks

Say that B_w is a γ***-smoothing network*** [1, 5] for some integer $\gamma \geq 1$ if for each input vector $\mathbf{x}$, $\mathsf{B}_w(\mathbf{x})$ is γ-smooth; so, the difference between the maximum and the minimum entries in the output vector of a γ-smoothing network is no more than γ. The ***smoothness*** of the network B_w is the least integer γ such that B_w is a γ-smoothing network.

For a number $\gamma > 0$, say that the network B_w is ***path*** γ***-concentrating on the input vector*** $\mathbf{x}$ if for each path $\beta \in \mathsf{P}[\mathsf{B}_w]$, there is a layer ℓ with $1 \leq \ell \leq |\beta|$ such that

$$\left| \widehat{x}(\mathsf{b}_r(\beta)) - \frac{\sum \mathbf{x}}{w} \right| \;\; \leq \;\; \gamma\,;$$

intuitively, on the input vector $\mathbf{x}$, every path from an input wire to an output wire in B_w includes a balancer b whose output $\widehat{x}(\mathsf{b})$ is within γ of the average $\frac{\sum \mathbf{x}}{w}$. Say that the network B_w is ***path*** γ***-concentrating*** if it is path γ-concentrating on every input vector $\mathbf{x}$.

3.3.2 A Combinatorial Lemma

We prove:

Lemma 3.2 *Consider a complete balancing network* B_w*. For some number* $\gamma \geq 2$*, assume that the network* B_w *is path* $(\gamma - 2)$*-concentrating. Fix a ceiling* γ*-concentrated input vector* $\mathbf{x}$*. Then, the output vector* $\mathbf{y}$ *is ceiling* $(\gamma - 1)$*-concentrated.*

Proof: Assume, by way of contradiction, that there is a wire $k_0 \in [w]$ such that

$$y_{k_0} \geq \left\lceil \frac{\sum \mathbf{x}}{w} \right\rceil + \gamma \,.$$

Since $\mathbf{x}$ is ceiling γ-concentrated, it follows that $\mathbf{y}$ is ceiling γ-concentrated. Hence, it follows that

$$y_{k_0} = \left\lceil \frac{\sum \mathbf{x}}{w} \right\rceil + \gamma \,.$$

We shall derive a contradiction to the assumption that B_w is path $(\gamma - 2)$-concentrating. specifically, we will prove that there is a path $\beta \in \mathsf{P}[\mathsf{B}_w]$ such that for every layer ℓ with $1 \leq \ell \leq \mathsf{d}(\mathsf{B}_w)$,

$$x_{i_1(\mathsf{b})}(\ell) = \left\lceil \frac{\sum \mathbf{x}}{w} \right\rceil + \gamma$$

and

$$x_{i_2(\mathsf{b})}(\ell) \geq \left\lceil \frac{\sum \mathbf{x}}{w} \right\rceil + \gamma - 1 \,.$$

We construct the path β by backward induction. For the sake of shortening the construction and the proof that the constructed path has the required property, we merge the basis case (where $\ell = \mathsf{d}(\mathsf{B}_w)$) and the induction step; so, the case $\ell = \mathsf{d}(\mathsf{B}_w)$ will be treated separately (where needed) along the induction step.

We assume, as our induction hypothesis, that we have defined a balancer b_r in layer ℓ with $1 < \ell \leq \mathsf{d}(\mathsf{B}_w)$ such that

$$x_{i_1(\mathsf{b}_r)}(\ell) = \left\lceil \frac{\sum \mathbf{x}}{w} \right\rceil + \gamma$$

and

$$x_{i_2(\mathsf{b}_r)}(\ell) \geq \left\lceil \frac{\sum \mathbf{x}}{w} \right\rceil + \gamma - 1 \,.$$

To ground the induction hypothesis, choose the balancer $\mathsf{b} = \mathsf{b}_{\mathsf{d}(\mathsf{B}_w)}$ so that $k_0 \in \{i_1(\mathsf{b}), i_2(\mathsf{b})\}$. (Since the network B_w is complete, there is such a balancer.) Hence, by assumption and the definition of balancer, it follows that

$$x_{i_1(\mathsf{b})} \quad = \quad \left\lceil \frac{\sum \mathbf{x}}{w} \right\rceil + \gamma$$

and

$$x_{i_2(\mathsf{b})} \quad \geq \quad \left\lceil \frac{\sum \mathbf{x}}{w} \right\rceil + \gamma - 1\,.$$

Since the network B_w is complete, there is a balancer $\mathsf{b} = \mathsf{b}_\ell$ in layer ℓ such that $i_1(\mathsf{b}_{\ell+1}) \in \{i_1(\mathsf{b}), i_2(\mathsf{b})\}$. Assume, without loss of generality, that $x_{i_1(\mathsf{b})}(\ell) \geq x_{i_2(\mathsf{b})}(\ell)$. Since $\mathbf{x}$ is ceiling γ-concentrated, it follows that $\mathbf{x}(\ell)$ is ceiling γ-concentrated. Hence, by the definition of balancer, it follows that

$$x_{i_2(\mathsf{b})}(\ell) \quad = \quad \left\lceil \frac{\sum \mathbf{x}}{w} \right\rceil + \gamma$$

and

$$x_{i_1(\mathsf{b})}(\ell) \quad \geq \quad \left\lceil \frac{\sum \mathbf{x}}{w} \right\rceil + \gamma - 1\,.$$

The inductive construction is complete, and the contradiction follows. ■

3.4 Randomized Balancing Networks

3.4.1 Randomized Balancers

A ***randomized balancer*** b [2, 13] is initialized to each of top and bottom with probability $\frac{1}{2}$ and independently of all other balancers; so, it is oriented uniformly at random.

Define a random variable r_{b} taking values $\frac{1}{2}$ and $-\frac{1}{2}$ with equal probability (cf. [13]). (Clearly, $\mathbb{E}\left[r_{\mathsf{b}}\right] = 0$.) Define also the random variable $\chi_{\mathsf{b}} = \mathsf{Odd}(\mathbf{x}) \cdot r_{\mathsf{b}}$ (cf. [13]). Then,

$$\begin{aligned} y_2 \quad &= \quad \frac{x_1 + x_2}{2} + \chi_{\mathsf{b}} \\ &= \quad \frac{x_1 + x_2}{2} + \mathsf{Odd}(\mathbf{x}_{\mathsf{b}}) \cdot r_{\mathsf{b}}. \end{aligned}$$

and

$$\begin{aligned} y_2 \quad &= \quad \frac{x_1 + x_2}{2} - \chi_{\mathsf{b}} \\ &= \quad \frac{x_1 + x_2}{2} - \mathsf{Odd}(\mathbf{x}_{\mathsf{b}}) \cdot r_{\mathsf{b}}. \end{aligned}$$

3.4.2 Randomized Networks

A ***randomized balancing network,*** or ***randomized network*** for short, consists of randomized balancers. So, a randomized balancing network is a balancing network B_w with a random orientation; it is represented by the set of (independent) random variables $\{r_{\mathsf{b}} \mid \mathsf{b} \in \mathcal{B}_w\}$. Note that each orientation of the network B_w determines values for all random variables $\{r_{\mathsf{b}} \mid \mathcal{B}_w\}$. So, a randomized network B_w induces a family of deterministic networks B_w, one for each possible orientation. An input vector $\mathbf{x}$ is a ***fixed-point*** of the randomized network w if $\mathbf{x}$ is a fixed-point for each induced deterministic network B_w.

Given an input vector $\mathbf{x}$ to a randomized balancing network B_w, induced in the natural way is a probability measure $\mathbb{P}$ on associated events. For each layer ℓ with $1 \leq \ell \leq \mathsf{d}(\mathsf{B}_w)$, induced is the random variable $\mathbf{y}(\ell)$, also called a *random vector*; clearly, the input vector to a restriction of the randomized network B_w is a random vector, and it will be called a *random input vector.* Note that for each wire $j \in [w]$ in layer ℓ, the random variable $y_j(\ell)$ is determined by *(i)* the inputs on each input wire $i \in \mathsf{ID}_j[1]$ and *(ii)* all random variables r_{b} where $\mathsf{b} \in \mathsf{D}_j[1, \ell]$. For some integer $y \geq 0$, say that the random variable $y_j(\ell)$ with $j \in [w]$ and $1 \leq \ell \leq \mathsf{d}(\mathsf{B}_w)$ ***attains the value*** y if there are values for the random variables $\{r_{\mathsf{b}} \mid \mathsf{b} \in \mathsf{D}_j[1, \ell]\}$ such that $y_j(\ell) = y$. Induced are also the sets of random variables $\{\mathsf{Odd}(\mathsf{x}_{\mathsf{b}}) \mid \mathsf{b} \in \mathcal{B}_w\}$ and $\{\chi_{\mathsf{b}} \mid \mathsf{b} \in \mathcal{B}_w\} = \{\mathsf{Odd}(\mathsf{x}_{\mathsf{b}}) \cdot r_{\mathsf{b}} \mid \mathsf{b} \in \mathcal{B}_w\}$; note that both are sets of not necessarily independent random variables.

We observe an immediate consequence of the disjointness among dependency sets estabished in Claim 3.1 for randomized bidelta networks:

Lemma 3.3 *Fix an input vector* $\mathbf{x}$ *to a bidelta network* B_w*. Then, for each path* $\beta \in \mathsf{P}[\mathsf{B}_w]$*, the set* $\bigcup_{\mathsf{b} \in \beta} \{\hat{x}(\mathsf{b})\}$ *is a set of independent random variables.*

Note that for a wire j and a layer ℓ in a complete balancing network B_w, the definition of a balancer implies that

$$y_j(\ell) \;=\; \frac{1}{2} \sum_{i' \in \mathsf{ID}_i[\ell-1]} y_{i'}(\ell-1) + \sum_{b \in \mathsf{D}_i[\ell-1]} \chi_{\mathsf{b}} \,.$$

We continue with an elementary observation on the relation between certain concentration and path concentration probabilities in a randomized complete balancing network.

Lemma 3.4 *Consider a randomized complete balancing network* $\mathsf{B}_w : \mathbf{x} \rightarrow \mathbf{y}$*, where* $\mathbf{x}$ *is a random input vector. Then,*

$$\begin{aligned} & \mathbb{P}\left[\mathbf{y} \textit{ is ceiling } (\gamma-1)\textit{-concentrated}\right] \\ \geq \;& \mathbb{P}\left[\mathbf{x} \textit{ is ceiling } \gamma\textit{-concentrated}\right] - \mathbb{P}\left[\mathsf{B}_w \textit{ is not path } (\gamma-2)\textit{-concentrating on } \mathbf{x}\right] . \end{aligned}$$

Proof: By Lemma 3.2 and the Union Bound (Lemma 2.3), we obtain that

$$\begin{aligned}
& \mathbb{P}\left[\mathbf{y} \text{ is ceiling } (\gamma-1)\text{-concentrated}\right] \\
\geq\ & \mathbb{P}\left[(\mathbf{x} \text{ is ceiling } \gamma\text{-concentrated}) \wedge (\mathsf{B}_w \text{ is path } (\gamma-2)\text{-concentrating on } \mathbf{x})\right] \\
=\ & 1-\mathbb{P}\left[(\mathbf{x} \text{ is not ceiling } \gamma\text{-concentrated}) \vee (\mathsf{B}_w \text{ is not path } (\gamma-2)\text{-concentrating on } \mathbf{x})\right] \\
\geq\ & 1-\mathbb{P}\left[\mathbf{x} \text{ is not ceiling } \gamma\text{-concentrated}\right]-\mathbb{P}\left[\mathsf{B}_w \text{ is not path } (\gamma-2)\text{-concentrating on } \mathbf{x}\right] \\
=\ & \mathbb{P}\left[\mathbf{x} \text{ is ceiling } \gamma\text{-concentrated}\right]-\mathbb{P}\left[\mathsf{B}_w \text{ is not path } (\gamma-2)\text{-concentrating on } \mathbf{x}\right],
\end{aligned}$$

as needed. ■

The next observation is an immediate consequence of the Law of Conditional Alternatives.

Claim 3.5 *Consider a randomized balancing network B_w with a random input vector $\mathbf{x}$. Then, for any event $\mathcal{E}$,*

$$\begin{aligned}
& \mathbb{P}\left[\mathsf{B}_w \textit{ is not path } \gamma\textit{-concentrating on } \mathbf{x}\right] \\
\leq\ & \mathbb{P}\left[\mathsf{B}_w \textit{ is not path } \gamma\textit{-concentrating on } \mathbf{x} \,\middle|\, \mathcal{E}\right]+\mathbb{P}\left[\neg\mathcal{E}\right].
\end{aligned}$$

3.4.3 Randomized Smoothing Networks

For some integer $\gamma \geq 1$, say that B_w is a ***γ-smoothing network with probability*** δ, where $0 \leq \delta \leq 1$, if for each input vector $\mathbf{x}$,

$$\mathbb{P}\left[\mathsf{B}_w(\mathbf{x}) \text{ is } \gamma\text{-smooth}\right] \quad \geq \quad \delta;$$

that is, the probability that for each pair of output wires $j, k \in [w]$, $|y_j - y_k| \leq \gamma$ is at least δ.

We observe:

Lemma 3.6 *Consider two topologically equivalent randomized networks. Then, their smoothnesses are identically distributed random variables.*

3.4.4 A Combinatorial Expression

We provide a combinatorial expression for the output vector of a (randomized) complete balancing network B_w in terms of the output vector of an earlier layer.

Lemma 3.7 *Consider a complete balancing network B_w. Then, for each pair of an output wire $j \in [w]$ and a layer ℓ with $0 \leq \ell < \mathsf{d}(\mathsf{B}_w)$,*

$$y_j \quad = \quad \frac{1}{2^{\mathsf{d}(\mathsf{B}_w)-\ell}} \sum_{i \in \mathsf{ID}_j[\ell]} y_i(\ell) + \sum_{k=\ell+1}^{\mathsf{d}(\mathsf{B}_w)} \frac{1}{2^{\mathsf{d}(\mathsf{B}_w)-k}} \sum_{\mathsf{b} \in \mathsf{D}_j[k]} \chi(\mathsf{b}).$$

Proof: By backward induction on ℓ. For the basis case, $\ell = \lg w - 1$. By the definition of balancer,

$$\begin{aligned} y_j &= \frac{1}{2} \sum_{i \in \mathsf{ID}_j[\mathsf{d}(\mathsf{B}_w)-1]} y_i(\mathsf{d}(\mathsf{B}_w)-1) + \sum_{\mathsf{b} \in \mathsf{D}_j[\mathsf{d}(\mathsf{B}_w)]} \chi_{\mathsf{b}} \\ &= \frac{1}{2^{\mathsf{d}(\mathsf{B}_w)-(\mathsf{d}(\mathsf{B}_w)-1)}} \sum_{i \in \mathsf{ID}_j[\mathsf{d}(\mathsf{B}_w)-1]} y_i(\mathsf{d}(\mathsf{B}_w)-1) + \sum_{k=(\mathsf{d}(\mathsf{B}_w)-1)+1}^{\mathsf{d}(\mathsf{B}_w)} \frac{1}{2^{\mathsf{d}(\mathsf{B}_w)-k}} \sum_{\mathsf{b} \in \mathsf{D}_j[k]} \chi_{\mathsf{b}}\,, \end{aligned}$$

and the claim follows.

Assume inductively that the claim holds for layer ℓ where $0 < \ell < \mathsf{d}(\mathsf{B}_w)$, and consider layer $\ell - 1$. By induction hypothesis and the definition of balancer,

$$\begin{aligned} y_j &= \frac{1}{2^{\mathsf{d}(\mathsf{B}_w)-\ell}} \sum_{i \in \mathsf{ID}_j[\ell]} y_i(\ell) + \sum_{k=\ell+1}^{\mathsf{d}(\mathsf{B}_w)} \frac{1}{2^{\mathsf{d}(\mathsf{B}_w)-k}} \sum_{b \in \mathsf{D}_j[k]} \chi_{\mathsf{b}} \\ &= \frac{1}{2^{\mathsf{d}(\mathsf{B}_w)-\ell}} \sum_{i \in \mathsf{ID}_j[\ell]} \left(\frac{1}{2} \sum_{i' \in \mathsf{ID}_i[\ell-1]} y_{i'}(\ell-1) + \sum_{b \in \mathsf{D}_i[\ell-1]} \chi_{\mathsf{b}} \right) + \sum_{k=\ell+1}^{\mathsf{d}(\mathsf{B}_w)} \frac{1}{2^{\mathsf{d}(\mathsf{B}_w)-k}} \sum_{b \in \mathsf{D}_j[k]} \chi_{\mathsf{b}} \\ &= \frac{1}{2^{\mathsf{d}(\mathsf{B}_w)-(\ell-1)}} \sum_{i' \in \mathsf{ID}_j[\ell-1]} y_{i'}(\ell-1) + \sum_{k=(\ell-1)+1}^{\mathsf{d}(\mathsf{B}_w)} \frac{1}{2^{\mathsf{d}(\mathsf{B}_w)-k}} \sum_{b \in \mathsf{D}_j[k]} \chi_{\mathsf{b}}\,, \end{aligned}$$

as needed. ■

We remark that Lemma 3.7 is motivated by and generalizes a corresponding equation derived in [13, Section 2.1] for the particular case where $\ell = 0$ and B_w is the block network [9].

3.4.5 A Probabilistic Lemma

Consider a collection of balancers $\mathcal{B}$ in a network B_w, with a constant c_{b} specific to each balancer $\mathsf{b} \in \mathcal{B}$. Following the notation from [13, p. 15], define the random variables

$$\mathsf{W}(\mathcal{B}) = \sum_{\mathsf{b} \in \mathcal{B}} c_{\mathsf{b}}\, r_{\mathsf{b}}\,,$$

$$\begin{aligned} \mathsf{V}(\mathcal{B}) &= \sum_{\mathsf{b} \in \mathcal{B}} c_{\mathsf{b}}\, \chi_{\mathsf{b}} \\ &= \sum_{\mathsf{b} \in \mathcal{B} \,|\, \mathsf{Odd}(\mathsf{x}_{\mathsf{b}})=1} c_{\mathsf{b}}\, r_{\mathsf{b}}\,, \end{aligned}$$

and

$$\begin{aligned} \mathsf{U}(\mathcal{B}) &= \mathsf{W}(\mathcal{B}) - \mathsf{V}(\mathcal{B}) \\ &= \sum_{\mathsf{b} \in \mathcal{B} \,|\, \mathsf{Odd}(\mathsf{x}_{\mathsf{b}})=0} c_{\mathsf{b}}\, r_{\mathsf{b}}\,. \end{aligned}$$

Clearly, the summation domain for the variable $\mathsf{W}(\mathcal{B})$ is *fixed,* so that $\mathsf{W}(\mathcal{B})$ is a fixed-domain sum of independent random variables; since for each random variable r_{b} with $\mathsf{b} \in \mathcal{B}$, r_{b} is symmetrically distributed around 0 (with $\mathbb{E}[r_{\mathsf{b}}] = 0$), linearity of expectation implies that $\mathbb{E}[\mathsf{W}(\mathcal{B})] = 0$, so that $\mathsf{W}(\mathcal{B})$ is symmetrically distributed around 0. This implies that

$$\mathbb{P}[\mathsf{W}(\mathcal{B}) < 0] \quad \leq \quad \frac{1}{2},$$

and

$$\mathbb{P}[\mathsf{W}(\mathcal{B}) > 0] \quad \leq \quad \frac{1}{2}.$$

In contrast, each of the summation domains $\{\mathsf{b} \in \mathcal{B} \mid \mathsf{Odd}(\mathbf{x}) = 1\}$ and $\{\mathsf{b} \in \mathcal{B} \mid \mathsf{Odd}(\mathbf{x}) = 0\}$ for $\mathsf{V}(\mathcal{B})$ and $\mathsf{U}(\mathcal{B})$, respectively, is a random variable itself; so, each of $\mathsf{V}(\mathcal{B})$ and $\mathsf{U}(\mathcal{B})$ is a non-fixed-domain sum of (not necessarily independent) random variables.

We continue with a formal proof for a (variant of) a claim from Herlihy and Tirthapura [13, Lemma 7].

Lemma 3.8 ([13]) *Consider a collection of balancers $\mathcal{B}$ in the randomized network B_w. Then, for any number $\delta > 0$,*

$$\mathbb{P}[|\mathsf{V}(\mathcal{B})| \geq \delta] \quad \leq \quad 2 \cdot \mathbb{P}[|\mathsf{W}(\mathcal{B})| \geq \delta] \ .$$

Proof: Clearly, by the definition of conditional probability,

$$\begin{aligned}
& \mathbb{P}[|\mathsf{W}(\mathcal{B})| \geq \delta] \\
= \;& \mathbb{P}[(\mathsf{W}(\mathcal{B}) \geq \delta) \vee (\mathsf{W}(\mathcal{B}) \leq -\delta)] \\
= \;& \mathbb{P}[\mathsf{W}(\mathcal{B}) \geq \delta] + \mathbb{P}[\mathsf{W}(\mathcal{B}) \leq -\delta] \\
\geq \;& \mathbb{P}[\mathsf{W}(\mathcal{B}) \geq \delta \wedge \mathsf{V}(\mathcal{B}) \geq \delta] + \mathbb{P}[\mathsf{W}(\mathcal{B}) \leq -\delta \wedge \mathsf{V}(\mathcal{B}) \leq -\delta] \\
= \;& \mathbb{P}[\mathsf{W}(\mathcal{B}) \geq \delta \mid \mathsf{V}(\mathcal{B}) \geq \delta] \cdot \mathbb{P}[\mathsf{V}(\mathcal{B}) \geq \delta] + \mathbb{P}[\mathsf{W}(\mathcal{B}) \leq -\delta \mid \mathsf{V}(\mathcal{B}) \leq -\delta] \cdot \mathbb{P}[\mathsf{V}(\mathcal{B}) \leq -\delta] \\
= \;& (1 - \mathbb{P}[\mathsf{W}(\mathcal{B}) < \delta \mid \mathsf{V}(\mathcal{B}) \geq \delta]) \cdot \mathbb{P}[\mathsf{V}(\mathcal{B}) \geq \delta] + (1 - \mathbb{P}[\mathsf{W}(\mathcal{B}) > -\delta \mid \mathsf{V}(\mathcal{B}) \leq -\delta]) \cdot \mathbb{P}[\mathsf{V}(\mathcal{B}) \leq -\delta] \\
= \;& \mathbb{P}[\mathsf{V}(\mathcal{B}) \geq \delta] + \mathbb{P}[\mathsf{V}(\mathcal{B}) \leq -\delta] \\
& -\mathbb{P}[\mathsf{W}(\mathcal{B}) < \delta \mid \mathsf{V}(\mathcal{B}) \geq \delta] \cdot \mathbb{P}[\mathsf{V}(\mathcal{B}) \geq \delta] - \mathbb{P}[\mathsf{W}(\mathcal{B}) > -\delta \mid \mathsf{V}(\mathcal{B}) \leq -\delta] \cdot \mathbb{P}[\mathsf{V}(\mathcal{B}) \leq -\delta] \\
= \;& \mathbb{P}[|\mathsf{V}(\mathcal{B})| \geq \delta] \\
& -\mathbb{P}[\mathsf{W}(\mathcal{B}) < \delta \mid \mathsf{V}(\mathcal{B}) \geq \delta] \cdot \mathbb{P}[\mathsf{V}(\mathcal{B}) \geq \delta] - \mathbb{P}[\mathsf{W}(\mathcal{B}) > -\delta \mid \mathsf{V}(\mathcal{B}) \leq -\delta] \cdot \mathbb{P}[\mathsf{V}(\mathcal{B}) \leq -\delta] \ .
\end{aligned}$$

Conditioned on the event $(\mathsf{V}(\mathcal{B}) \geq \delta)$ (resp., the event $(\mathsf{V}(\mathcal{B}) \leq -\delta)$), the event $(\mathsf{W}(\mathcal{B}) < \delta)$ (resp., the event $(\mathsf{W}(\mathcal{B}) > -\delta)$) implies the event $(\mathsf{U}(\mathcal{B}) < 0)$ (resp., the event $(\mathsf{U}(\mathcal{B}) > 0)$).

Hence,

$$
\begin{aligned}
& \mathbb{P}\left[|\mathsf{W}(\mathcal{B})| \geq \delta\right] \\
\geq\ & \mathbb{P}\left[|\mathsf{V}(\mathcal{B})| \geq \delta\right] \\
& -\mathbb{P}\left[\mathsf{U}(\mathcal{B}) < 0 \mid \mathsf{V}(\mathcal{B}) \geq \delta\right] \cdot \mathbb{P}\left[\mathsf{V}(\mathcal{B}) \geq \delta\right] - \mathbb{P}\left[\mathsf{U}(\mathcal{B}) > 0 \mid \mathsf{V}(\mathcal{B}) \leq -\delta\right] \cdot \mathbb{P}\left[\mathsf{V}(\mathcal{B}) \leq -\delta\right] \\
=\ & \mathbb{P}\left[|\mathsf{V}(\mathcal{B})| \geq \delta\right] - \mathbb{P}\left[(\mathsf{U}(\mathcal{B}) < 0) \wedge (\mathsf{V}(\mathcal{B}) \geq \delta)\right] - \mathbb{P}\left[(\mathsf{U}(\mathcal{B}) > 0) \wedge (\mathsf{V}(\mathcal{B}) \leq -\delta)\right] \\
=\ & \mathbb{P}\left[|\mathsf{V}(\mathcal{B})| \geq \delta\right] \\
& - \sum_{\mathcal{B}' \subseteq \mathcal{B}} \mathbb{P}\left[\{\mathsf{b} \mid \mathsf{Odd}(\mathbf{x}_\mathsf{b}) = 0\} = \mathcal{B}'\right] \cdot \mathbb{P}\left[(\mathsf{U}(\mathcal{B}) < 0) \wedge (\mathsf{V}(\mathcal{B}) \geq \delta) \mid \{\mathsf{b} \mid \mathsf{Odd}(\mathbf{x}_\mathsf{b}) = 0\} = \mathcal{B}'\right] \\
& - \sum_{\mathcal{B}' \subseteq \mathcal{B}} \mathbb{P}\left[\{\mathsf{b} \mid \mathsf{Odd}(\mathbf{x}_\mathsf{b}) = 0\} = \mathcal{B}'\right] \cdot \mathbb{P}\left[(\mathsf{U}(\mathcal{B}) > 0) \wedge (\mathsf{V}(\mathcal{B}) \leq -\delta) \mid \{\mathsf{b} \mid \mathsf{Odd}(\mathbf{x}_\mathsf{b}) = 0\} = \mathcal{B}'\right].
\end{aligned}
$$

We prove:

Claim 3.9 *Fix a set of balancers $\mathcal{B}' \subseteq \mathcal{B}$. Then, the events $(\mathsf{U}(\mathcal{B}) < 0)$ and $(\mathsf{V}(\mathcal{B}) \geq \delta)$ (resp., the events $(\mathsf{U}(\mathcal{B}) > 0)$ and $(\mathsf{V}(\mathcal{B}) \leq -\delta)$) are conditionally independent given the event $\{\mathsf{b} \mid \mathsf{Odd}(\mathbf{x}_\mathsf{b}) = 0\} = \mathcal{B}'$.*

Proof: Since $\{\mathsf{b} \mid \mathsf{Odd}(\mathbf{x}_\mathsf{b}) = 0\} = \mathcal{B}'$, the summation domains for $\mathsf{U}(\mathcal{B})$ and $\mathsf{V}(\mathcal{B})$ are fixed (and not random variables); since they are disjoint, it follows that the random variables $\mathsf{U}(\mathcal{B})$ and $\mathsf{V}(\mathcal{B})$ are conditionally independent (given the event $\{\mathsf{b} \mid \mathsf{Odd}(\mathbf{x}_\mathsf{b}) = 0\} = \mathcal{B}'$). This implies that the events $(\mathsf{U}(\mathcal{B}) < 0)$ and $(\mathsf{V}(\mathcal{B}) \geq \delta)$ (resp., the events $(\mathsf{U}(\mathcal{B}) > 0)$ and $(\mathsf{V}(\mathcal{B}) \leq -\delta)$) are also conditionally independent. ■

By Claim 3.9, it follows that

$$
\begin{aligned}
& \mathbb{P}\left[|\mathsf{W}(\mathcal{B})| \geq \delta\right] \\
\geq\ & \mathbb{P}\left[|\mathsf{V}(\mathcal{B})| \geq \delta\right] - \sum_{\mathcal{B}' \subseteq \mathcal{B}} \mathbb{P}\left[\{\mathsf{b} \mid \mathsf{Odd}(\mathbf{x}_{\mathsf{b}}) = 0\} = \mathcal{B}'\right] \cdot \\
& (\mathbb{P}\left[\mathsf{U}(\mathcal{B}) < 0 \mid \{\mathsf{b} \mid \mathsf{Odd}(\mathbf{x}_{\mathsf{b}}) = 0\} = \mathcal{B}'\right] \cdot \mathbb{P}\left[\mathsf{V}(\mathcal{B}) \geq \delta \mid \{\mathsf{b} \mid \mathsf{Odd}(\mathbf{x}_{\mathsf{b}}) = 0\} = \mathcal{B}'\right] \\
& + \mathbb{P}\left[\mathsf{U}(\mathcal{B}) > 0 \mid \{\mathsf{b} \mid \mathsf{Odd}(\mathbf{x}_{\mathsf{b}}) = 0\} = \mathcal{B}'\right] \cdot \mathbb{P}\left[\mathsf{V}(\mathcal{B}) \leq -\delta \mid \{\mathsf{b} \mid \mathsf{Odd}(\mathbf{x}_{\mathsf{b}}) = 0\} = \mathcal{B}'\right]) \\
=\ & \mathbb{P}\left[|\mathsf{V}(\mathcal{B})| \geq \delta\right] - \sum_{\mathcal{B}' \subseteq \mathcal{B}} \mathbb{P}\left[\{\mathsf{b} \mid \mathsf{Odd}(\mathbf{x}_{\mathsf{b}}) = 0\} = \mathcal{B}'\right] \cdot \\
& (\mathbb{P}\left[\mathsf{W}(\mathcal{B}') < 0\right] \cdot \mathbb{P}\left[\mathsf{V}(\mathcal{B}) \geq \delta \mid \{\mathsf{b} \mid \mathsf{Odd}(\mathbf{x}_{\mathsf{b}}) = 0\} = \mathcal{B}'\right] \\
& + \mathbb{P}\left[\mathsf{W}(\mathcal{B}') > 0\right] \cdot \mathbb{P}\left[\mathsf{V}(\mathcal{B}) \leq -\delta \mid \{\mathsf{b} \mid \mathsf{Odd}(\mathbf{x}_{\mathsf{b}}) = 0\} = \mathcal{B}'\right]) \\
\geq\ & \mathbb{P}\left[|\mathsf{V}(\mathcal{B})| \geq \delta\right] - \sum_{\mathcal{B}' \subseteq \mathcal{B}} \mathbb{P}\left[\{\mathsf{b} \mid \mathsf{Odd}(\mathbf{x}_{\mathsf{b}}) = 0\} = \mathcal{B}'\right] \cdot \\
& \left(\frac{1}{2} \cdot \mathbb{P}\left[\mathsf{V}(\mathcal{B}) \geq \delta \mid \{\mathsf{b} \mid \mathsf{Odd}(\mathbf{x}_{\mathsf{b}}) = 0\} = \mathcal{B}'\right] + \frac{1}{2} \cdot \mathbb{P}\left[\mathsf{V}(\mathcal{B}) \leq -\delta \mid \{\mathsf{b} \mid \mathsf{Odd}(\mathbf{x}_{\mathsf{b}}) = 0\} = \mathcal{B}'\right]\right) \\
=\ & \mathbb{P}\left[|\mathsf{V}(\mathcal{B})| \geq \delta\right] - \frac{1}{2} \sum_{\mathcal{B}' \subseteq \mathcal{B}} \mathbb{P}\left[\{\mathsf{b} \mid \mathsf{Odd}(\mathbf{x}_{\mathsf{b}}) = 0\} = \mathcal{B}'\right] \cdot \\
& (\mathbb{P}\left[\mathsf{V}(\mathcal{B}) \geq \delta \mid \{\mathsf{b} \mid \mathsf{Odd}(\mathbf{x}_{\mathsf{b}}) = 0\} = \mathcal{B}'\right] + \mathbb{P}\left[\mathsf{V}(\mathcal{B}) \leq -\delta \mid \{\mathsf{b} \mid \mathsf{Odd}(\mathbf{x}_{\mathsf{b}}) = 0\} = \mathcal{B}'\right]) \\
=\ & \mathbb{P}\left[|\mathsf{V}(\mathcal{B})| \geq \delta\right] - \frac{1}{2} \sum_{\mathcal{B}' \subseteq \mathcal{B}} \mathbb{P}\left[\{\mathsf{b} \mid \mathsf{Odd}(\mathbf{x}_{\mathsf{b}}) = 0\} = \mathcal{B}'\right] \cdot \\
& \mathbb{P}\left[(\mathsf{V}(\mathcal{B}) \geq \delta) \vee (\mathsf{V}(\mathcal{B}) \leq -\delta) \mid \{\mathsf{b} \mid \mathsf{Odd}(\mathbf{x}_{\mathsf{b}}) = 0\} = \mathcal{B}'\right] \\
=\ & \mathbb{P}\left[|\mathsf{V}(\mathcal{B})| \geq \delta\right] - \frac{1}{2} \sum_{\mathcal{B}' \subseteq \mathcal{B}} \mathbb{P}\left[\{\mathsf{b} \mid \mathsf{Odd}(\mathbf{x}_{\mathsf{b}}) = 0\} = \mathcal{B}'\right] \cdot \mathbb{P}\left[|\mathsf{V}(\mathcal{B})| \geq \delta \mid \{\mathsf{b} \mid \mathsf{Odd}(\mathbf{x}_{\mathsf{b}}) = 0\} = \mathcal{B}'\right] \\
=\ & \mathbb{P}\left[|\mathsf{V}(\mathcal{B})| \geq \delta\right] - \frac{1}{2} \mathbb{P}\left[|\mathsf{V}(\mathcal{B})| \geq \delta\right] \\
=\ & \frac{1}{2} \mathbb{P}\left[|\mathsf{V}(\mathcal{B})| \geq \delta\right],
\end{aligned}
$$

as needed. ■

4 Block Network (and Relatives)

Henceforth, denote as Block_w the ***block network*** of width w [9], where w is a power of 2. Section 4.1 reviews the inductive construction of the block network. Section 4.2 reviews the tree structure from [13, Section 2] associated with the block network. The randomized block network is treated in Section 4.3. Section 4.4 deals with the (relative) cube-connected-cycles network.

4.1 Inductive Construction of Block_w

- <u>*Basis:*</u> The network Block_2 is a single balancer.

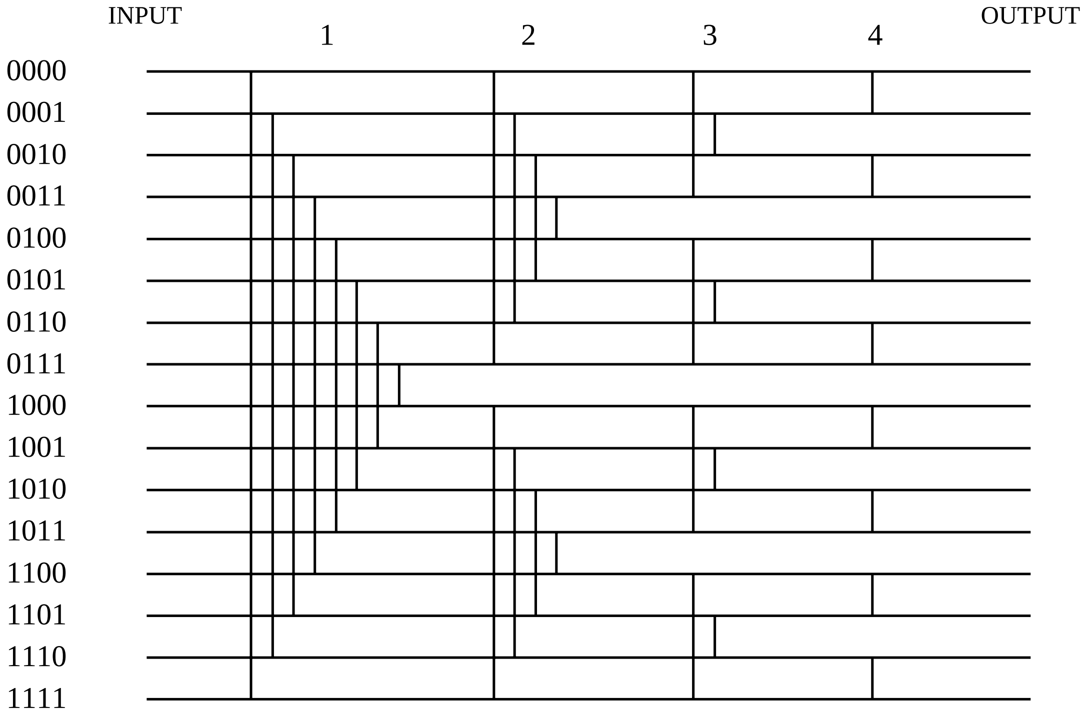

Figure 3: The network Block_{16}. The binary strings next to the input wires are their correponding binary representations. The four numbers on the top indicate the four layers.

- *Induction step:* The network Block_{2w} is constructed from two copies of the network Block_w and an additional layer as follows. For an input vector $\mathbf{x}$, each of $\mathbf{x}_A$ and $\mathbf{x}_B$ is fed into each of two copies of Block_w; denote as $\mathbf{y}_A$ and $\mathbf{y}_B$ the output vectors of the two copies. In a rightmost layer, each pair of corresponding entries of the vectors $\mathbf{y}_A$ and $\mathbf{y}_B$ are matched through a balancer. (See Figure 3 for an illustration.)

So, the network Block_w has $\lg w$ complete layers $1, \ldots, \lg w$, and $\mathsf{d}(\mathsf{Block_w}) = \lg \mathsf{w}$.

The ***periodic network*** of width w [9], denoted as $\mathsf{Periodic}_w$, is the cascade of $\lg w$ copies of the network Block_w; so, $\mathsf{d}(\mathsf{Periodic}_w) = \lg^2 w$.

4.2 The Tree Structure

The network Block_w induces a (rooted) binary tree, where each node is associated with a set of balancers from the same layer of the network; different nodes are associated with disjoint sets, and we shall often identify a node with the corresponding set of balancers. Each node is labeled with two integers; the first one is the layer of the associated balancers.

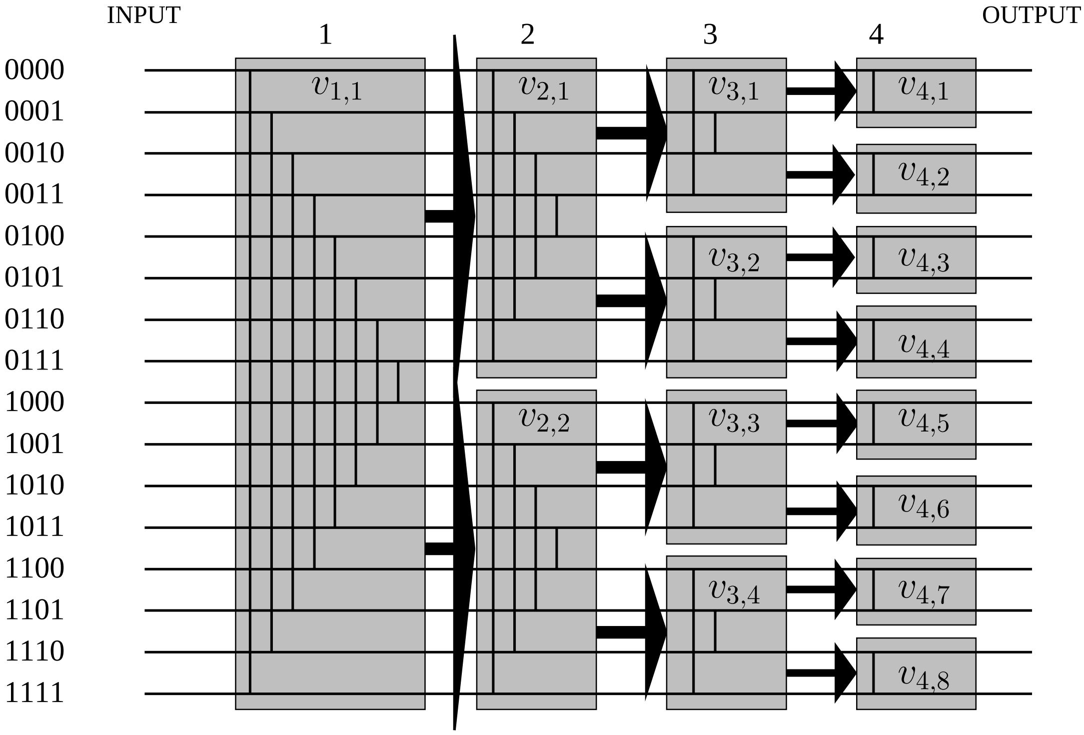

Figure 4: The tree structure for the network Block_{16}. Each node is shown as a shaded box.

- The *root,* denoted as $v_{1,1}$, consists of all $\frac{w}{2}$ balancers from layer 1.
- Each layer ℓ with $2 \leq \ell \leq \lg w$ induces $2^{\ell-1}$ nodes, denoted as $v_{\ell,1}, \ldots, v_{\ell,2^{\ell-1}}$ and each consisting of $\frac{w}{2^\ell}$ balancers. The nodes for layer ℓ are defined inductively given the nodes for layer $\ell - 1$. Fix an integer j with $1 \leq j \leq 2^{\ell-2}$.
  - The node $v_{\ell,2j-1}$ consists of all balancers (in layer ℓ) where the top output wire of some balancer in node $v_{\ell-1,j}$ is connected to.
  - The node $v_{\ell,2j}$ consists of all balancers (in layer ℓ) where the bottom output wire of some balancer in node $v_{\ell-1,j}$ is connected to.

See Figure 4 for an illustration of the tree structure. Clearly, the tokens exiting from output wire 0 of the network Block_w must have followed the path $v_{1,1}, \ldots, v_{\lg w,1}$ and exited on the top output wire of the single balancer in $v_{\lg w,1}$. We continue with a simple observation.

Claim 4.1 *For the network* Block_w*,*

$$\sum_{\ell=1}^{\lg w - \lceil \lg \lg w \rceil} \sum_{\mathsf{b} \in v_{\ell,1}} \left(\frac{1}{2^{\lg w - \ell + 1}} - \left(- \frac{1}{2^{\lg w - \ell + 1}} \right) \right)^2 \;\leq\; \frac{2}{\lg w} \,.$$

Proof: Clearly,

$$\begin{aligned}
\sum_{\ell=1}^{\lg w - \lceil \lg \lg w \rceil} \sum_{\mathsf{b} \in v_{\ell,1}} \left(\frac{1}{2^{\lg w - \ell + 1}} - \left(-\frac{1}{2^{\lg w - \ell + 1}} \right) \right)^2 &= \sum_{\ell=1}^{\lg w - \lceil \lg \lg w \rceil} 2^{\lg w - \ell} \left(\frac{1}{2^{\lg w - \ell}} \right)^2 \\
&= \sum_{\ell=1}^{\lg w - \lceil \lg \lg w \rceil} \frac{1}{2^{\lg w - \ell}} \\
&\leq \frac{1}{w} \cdot (2^{\lg w - \lg \lg w + 1} - 1) \\
&\leq \frac{2}{\lg w},
\end{aligned}$$

as needed. ■

For a layer ℓ and an integer j with $1 \leq j \leq 2^{\ell-1}$ denote as $x_{\ell,j}$ and $y_{\ell,j}$ the number of tokens entering and exiting all balancers in node $v_{\ell,j}$, respectively.

4.3 Randomized Block Network

The numbers of tokens $y_0, \ldots, y_{w-1}$ on the output wires $0, \ldots, w-1$, respectively, of the randomized block network are random variables. for each pair of a layer ℓ and an integer j with the numbers $x_{\ell,j}$ and $y_{\ell,j}$ are also random variables. Fix a layer $\ell > 1$; since the tokens exiting on the top (resp., bottom) output wires of balancers associated with node $v_{\ell-1,j}$ enter node $v_{\ell,2j-1}$ (resp., $v_{\ell,2j}$), it follows that

$$x_{\ell,2j-1} = \frac{x_{\ell-1,j}}{2} + \sum_{\mathsf{b} \in v_{\ell-1,j}} \chi_{\mathsf{b}}$$

and

$$x_{\ell,2j} = \frac{x_{\ell-1,j}}{2} - \sum_{\mathsf{b} \in v_{\ell-1,j}} \chi_{\mathsf{b}},$$

respectively.

The symmetry of the block network implies that all random variables y_j with $j \in [w]$ follow the same distribution (cf. [13, Proof of Theorem 10]); so, we shall only analyze y_0. wire of $v_{\lg w,1}$. We recall a preliminary claim due to Herlihy and Tirthapura [13].

Claim 4.2 ([13]) *For the randomized network* Block_w,

$$y_0 = \frac{\sum \mathbf{x}}{w} + \sum_{\ell=1}^{\lg w} \frac{1}{2^{\lg w - \ell}} \sum_{\mathsf{b} \in v_{\ell,1}} \chi_{\mathsf{b}}.$$

The following technical claim can be derived easily using techniques from [13].

Lemma 4.3 *For the randomized network* Block_w *and for any number* $\delta > 0$,

$$\mathbb{P}\left[\left|y_0 - \frac{\sum \mathbf{x}}{w}\right| \geq \delta\right] \quad \leq \quad 4 \cdot \mathsf{exp}(-\delta^2)\,.$$

The proof will use a simple calculation from [13, page 5]:

$$\sum_{\ell=1}^{\lg w} \sum_{\mathsf{b} \in v_{\ell,1}} \left(\frac{1}{2^{\lg w - \ell + 1}} - \left(-\frac{1}{2^{\lg w - \ell + 1}}\right)\right)^2 \quad = \quad 2 - \frac{2}{w}\,.$$

Proof: By Claim 4.2,

$$\mathbb{P}\left[\left|y_0 - \frac{\sum \mathbf{x}}{w}\right| \geq \delta\right] \quad = \quad \mathbb{P}\left[\left|\sum_{\ell=1}^{\lg w} \sum_{\mathsf{b} \in v_{\ell,1}} \frac{1}{2^{\lg w - \ell}} \chi_{\mathsf{b}}\right| \geq \delta\right]\,.$$

Now use Claim 3.8 with $\mathcal{B} = \bigcup_{1 \leq \ell \leq \lg w} v_{\ell,1}$ and $c_{\mathsf{b}} = \frac{1}{2^{\lg w - \ell}}$ for each balancer $\mathsf{b} \in v_{\ell,1}$ with $1 \leq \ell \leq \lg w$. Then,

$$\mathbb{P}\left[\left|\sum_{\ell=1}^{\lg w} \frac{1}{2^{\lg w - \ell}} \sum_{\mathsf{b} \in v_{\ell,1}} \chi_{\mathsf{b}}\right| \geq \delta\right] \quad \leq \quad 2 \cdot \mathbb{P}\left[\left|\sum_{\ell=1}^{\lg w} \frac{1}{2^{\lg w - \ell}} \sum_{\mathsf{b} \in v_{\ell,1}} r_{\mathsf{b}}\right| \geq \delta\right]\,.$$

Note that for each pair of a layer ℓ with $1 \leq \ell \leq \lg w$ and a balancer $\mathsf{b} \in v_{\ell,1}$, the random variable $\frac{1}{2^{\lg w - \ell}} \cdot r_{\mathsf{b}}$ has range $\left\{-\frac{1}{2^{\lg w - \ell + 1}}, +\frac{1}{2^{\lg w - \ell + 1}}\right\}$. By Hoeffding Bound (Lemma 2.4), it follows that

$$\begin{aligned}
\mathbb{P}\left[\left|y_0 - \frac{\sum \mathbf{x}}{w}\right| \geq \delta\right] \quad & \leq \quad 2 \cdot 2 \cdot \mathsf{exp}\left(-\frac{2\delta^2}{\sum_{\ell=1}^{\lg w} \sum_{\mathsf{b} \in v_{\ell,1}} \left(\frac{1}{2^{\lg w - \ell + 1}} - \left(-\frac{1}{2^{\lg w - \ell + 1}}\right)\right)^2}\right) \\
& = \quad 4 \cdot \mathsf{exp}\left(-\frac{2\delta^2}{2 - \frac{2}{w}}\right) \\
& \leq \quad 4 \cdot \mathsf{exp}(-\delta^2)\,,
\end{aligned}$$

as needed. ■

4.4 The Cube-Connected-Cycles Network

Henceforth, denote as CCC_w the ***cube-connected-cycles network*** [21] of width w, where w is a power of 2.

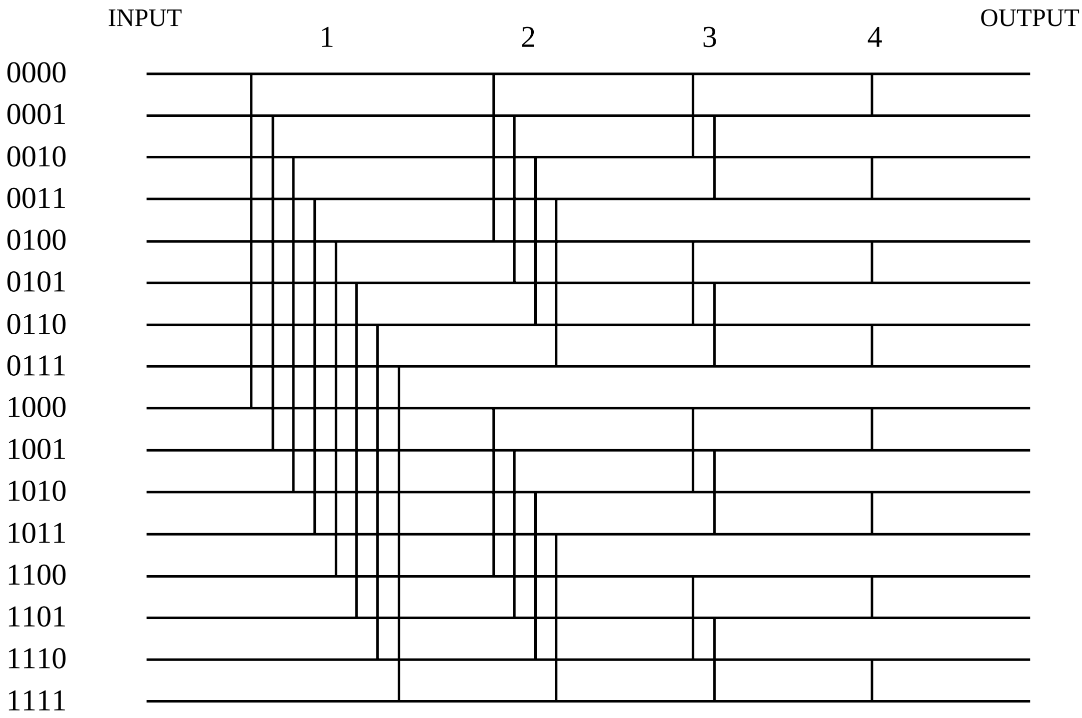

Figure 5: The network CCC_{16}.

4.4.1 Construction

The network CCC_w has $\lg w$ (complete) layers. For each layer ℓ with $1 \leq \ell \leq \lg w$, for each wire $u \in \{0,1\}^{\lg w}$, there is a balancer b connecting wire u and wire $u(\ell)$. (See Figure 5 for an illustration to the construction.) Clearly, the network CCC_w is the sequential cascade of the ladder network L_w followed by the parallel cascade of two copies of the network $\mathsf{CCC}_{\frac{w}{2}}$.

4.4.2 Topological Properties

We start a simple property of the cube-connected-cycles network.

Lemma 4.4 *Fix a pair of integers ℓ_1 and ℓ_2 with $\ell_1 + \ell_2 < \lg w$, and a corresponding pair of binary strings $l_1 \in \{0,1\}^{\ell_1}$ and $l_2 \in \{0,1\}^{\ell_2}$. Then, the network $\mathsf{CCC}_w \setminus [\ell_1, \ell_2]$ restricted to the set of wires $\left\{l_1 u l_2 \mid u \in \{0,1\}^{\lg w - \ell_1 - \ell_2}\right\}$ is a cube-connected-cycles network $\mathsf{CCC}_{2^{\lg w - \ell_1 - \ell_2}}$.*

Proof: We provide a balancer-preserving permutation $\pi : \left\{l_1 u l_2 \mid u \in \{0,1\}^{\lg w - \ell_1 - \ell_2}\right\} \to \{0,1\}^{\lg w - \ell_1 - \ell_2}$ between the sets of wires of the networks $\mathsf{CCC}_w \setminus [\ell_1, \ell_2]$ and $\mathsf{CCC}_{2^{\lg w - \ell_1 - \ell_2}}$, respectively: For an arbitrary wire $l_1 u l_2$ with $u \in \{0,1\}^{\lg w - \ell_1 - \ell_2}$, $\pi(l_1 u l_2) = u$.

Consider a balancer b in layer $\ell - \ell_1$ of the network $\mathsf{CCC}_w \setminus [\ell_1, \ell_2]$, with $\ell_1 + 1 \leq \ell \leq \lg w - \ell_2$. By construction of the network CCC_w, b connects the wires $l_1 u l_2$ and $l_1 u(\ell - \ell_1) l_2$ for some $u \in \{0,1\}^{\lg w - \ell_1 - \ell_2}$. By definition of π, $\pi(l_1 u l_2) = u$ and $\pi(l_1 u(\ell - \ell_1) l_2) = u(\ell - \ell_1)$. By construction of the network $\mathsf{CCC}_{2^{\lg w - \ell_1 - \ell_2}}$, the wires u and $u(\ell - \ell_1)$ are connected by a balancer in the network $\mathsf{CCC}_{2^{\lg w - \ell_1 - \ell_2}}$, and the claim follows. ■

By Lemma 4.4, it immediately follows:

Corollary 4.5 *Fix a pair of integers ℓ_1 and ℓ_2 with $\ell_1 + \ell_2 < \lg w$. Then, the network $\mathsf{CCC}_w \setminus [\ell_1, \ell_2]$ is the parallel cascade of $2^{\ell_1 + \ell_2}$ copies of the network $\mathsf{CCC}_{2^{\lg w - \ell_1 - \ell_2}}$.*

We continue with another simple property of the cube-connected-cycles network.

Lemma 4.6 *Consider layers ℓ and ℓ' in the network CCC_w^2, with $\lg w + 1 \leq \ell \leq 2 \lg w$ and $\ell - \lg w \leq \ell' < \ell$. Fix an input wire $i = i_1 \ldots i_{\lg w}$ in layer ℓ. Then,*

$$\mathsf{ID}_i(\ell') = \begin{cases} \left\{ u_1 i_{\ell - \lg w} \ldots i_{\ell' - 1} u_2 \mid u_1 \in \{0,1\}^{\ell - 1 - \lg w} \text{ and } u_2 \in \{0,1\}^{\lg w - \ell' + 1} \right\}, & \text{if } \ell' \leq \lg w \\ \left\{ i_1 \ldots i_{\ell' - \lg w - 1} u_1 i_{\ell - \lg w} \ldots i_{\lg w} \mid u_1 \in \{0,1\}^{\ell - \ell'} \right\}, & \text{if } \ell' > \lg w \end{cases}.$$

Proof: By backward induction on ℓ'. For the basis case, assume that $\ell' = \ell - 1$. We proceed by case analysis.

1. Assume first that $\ell' \leq \lg w$. Since $\ell \geq \lg w + 1$, it follows that $\ell' = \lg w$ and $\ell = \lg w + 1$. By construction of the network CCC_w, wire i is the output wire of a balancer in layer $\lg w$ (within the first cascaded network CCC_w) with input wires $i = i_1 \ldots i_{\lg w}$ and $i(\lg w) = i_1 \ldots \overline{i_{\lg w}}$. Hence,

$$\begin{aligned} \mathsf{ID}_i(\ell') &= \{ i_1 \ldots i_{\lg w - 1} u_2 \mid u_2 \in \{0,1\} \} \\ &= \left\{ i_{\ell - \lg w} \ldots i_{\ell' - 1} u_2 \mid u_2 \in \{0,1\}^{\lg w - \ell' + 1} \right\} \end{aligned}$$

   as needed.

2. Assume now that $\ell' > \lg w$. By construction of the network CCC_w, wire i is the output wire of a balancer in layer ℓ' (within the second cascaded network CCC_w) with input wires $i = i_1 \ldots i_{\ell' - \lg w - 1} i_{\ell' - \lg w} i_{\ell' - \lg w + 1} \ldots i_{\lg w}$ and $i(\ell' - \lg w) = i_1 \ldots i_{\ell' - \lg w - 1} \overline{i_{\ell' - \lg w}} i_{\ell' - \lg w + 1} \ldots i_{\lg w}$. Hence,

$$\begin{aligned} \mathsf{ID}_i(\ell') &= \{ i_1 \ldots i_{\ell' - \lg w - 1} u_1 i_{\ell' - \lg w + 1} \ldots i_{\lg w} \mid u_1 \in \{0,1\} \} \\ &= \{ i_1 \ldots i_{\ell' - \lg w - 1} u_1 i_{\ell - \lg w} \ldots i_{\lg w} \mid u_1 \in \{0,1\} \}, \end{aligned}$$

   as needed.

Assume inductively that the claim holds for some layer ℓ' with $\ell - \lg w < \ell' < \ell$. For the induction step, consider the integer $\ell' - 1$. We proceed by case analysis.

1. Assume first that $\ell' \leq \lg w$. By induction hypothesis,

$$\mathsf{ID}_i(\ell') \quad = \quad \left\{ u_1 i_{\ell-\lg w} \ldots i_{\ell'-1} u_2 \mid u_1 \in \{0,1\}^{\ell-1-\lg w} \text{ and } u_2 \in \{0,1\}^{\lg w - \ell' + 1} \right\} .$$

Fix some wire $j \in \mathsf{ID}_i(\ell')$; so, $j = u_1 i_{\ell-\lg w} \ldots i_{\ell'-1} u_2$ for a pair of binary strings $u_1 \in \{0,1\}^{\ell-1-\lg w}$ and $u_2 \in \{0,1\}^{\lg w-\ell'+1}$. By construction of the network CCC_w, wire j is the output wire of a balancer b in layer $\ell'-1$ (within the first cascaded network CCC_w). Then, b has input wires $j = u_1 i_{\ell-\lg w} \ldots i_{\ell'-2} i_{\ell'-1} u_2$ and $j(\ell'-1) = u_1 i_{\ell-\lg w} \ldots i_{\ell'-2} \overline{i_{\ell'-1}} u_2$. Hence,

$$\begin{aligned} \mathsf{ID}_i(\ell'-1) \quad &= \quad \left\{ u_1 i_{\ell-\lg w} \ldots i_{\ell'-2} u u_2 \mid u_1 \in \{0,1\}^{\ell-1-\lg w}, u \in \{0,1\} \text{ and } u_2 \in \{0,1\}^{\lg w-\ell'+1} \right\} \\ &= \quad \left\{ u_1 i_{\ell-\lg w} \ldots i_{\ell'-2} u_2 \mid u_1 \in \{0,1\}^{\ell-1-\lg w} \text{ and } u_2 \in \{0,1\}^{\lg w-(\ell'-1)+1} \right\} , \end{aligned}$$

as needed.

2. Assume now that $\ell' > \lg w$. By induction hypothesis,

$$\mathsf{ID}_i(\ell') \quad = \quad \left\{ i_1 \ldots i_{\ell'-\lg w-1} u_1 i_{\ell-\lg w} \ldots i_{\lg w} \mid u_1 \in \{0,1\}^{\ell-\ell'} \right\} .$$

Fix some wire $j \in \mathsf{ID}_i(\ell')$; so, $j = i_1 \ldots i_{\ell'-\lg w-1} u_1 i_{\ell-\lg w} \ldots i_{\lg w}$ for some binary string $u_1 \in \{0,1\}^{\ell-\ell'}$. By construction of the network CCC_w, wire j is the output wire of a balancer b in layer $\ell'-1$. We proceed by case analysis.

   (a) Assume first that $\ell'-1 > \lg w$, so that balancer b is within the second cascaded network CCC_w. Then, b has input wires $j = i_1 \ldots i_{\ell'-\lg w-2} i_{\ell'-\lg w-1} u_1 \ldots i_{\ell-\lg w} \ldots i_{\lg w}$ and $j(\ell'-\lg w-1) = i_1 \ldots i_{\ell'-\lg w-2} \overline{i_{\ell'-\lg w-1}} u_1 \ldots i_{\ell-\lg w} \ldots i_{\lg w}$. Hence,

$$\begin{aligned} \mathsf{ID}_i(\ell'-1) \quad &= \quad \left\{ i_1 \ldots i_{\ell'-\lg w-2} u u_1 i_{\ell-\lg w} \ldots i_{\lg w} \mid u \in \{0,1\} \text{ and } u_1 \in \{0,1\}^{\ell-\ell'} \right\} \\ &= \quad \left\{ i_1 \ldots i_{\ell'-\lg w-2} u_1 i_{\ell-\lg w} \ldots i_{\lg w} \mid u_1 \in \{0,1\}^{\ell-(\ell'-1)} \right\} , \end{aligned}$$

   as needed.

   (b) Assume now that $\ell'-1 \leq \lg w$, so that balancer b is within the first cascaded network CCC_w. It follows that $\ell' = \lg w + 1$. So,

$$\mathsf{ID}_i(\ell') \quad = \quad \left\{ u_1 i_{\ell-\lg w} \ldots i_{\lg w} \mid u_1 \in \{0,1\}^{\ell-\ell'} \right\} ,$$

   and $j = u_1 i_{\ell-\lg w} \ldots i_{\lg w}$. Then, b has input wires $j = u_1 i_{\ell-\lg w} \ldots i_{\lg w}$ and $j(\lg w) = u_1 i_{\ell-\lg w} \ldots \overline{i_{\lg w}}$. Hence,

$$\begin{aligned} \mathsf{ID}_i(\ell'-1) \quad &= \quad \left\{ u_1 i_{\ell-\lg w} \ldots i_{\lg w} u_2 \mid u_1 \in \{0,1\}^{\ell-\ell'} \text{ and } u_2 \in \{0,1\}^1 \right\} \\ &= \quad \left\{ u_1 i_{\ell-\lg w} \ldots i_{\lg w} u_2 \mid u_1 \in \{0,1\}^{\ell-1-\lg w} \text{ and } u_2 \in \{0,1\}^{\lg w-(\ell'-1)+1} \right\} , \end{aligned}$$

   as needed.

The induction is now complete. ■

It is simple to see that the cube-connected-cycles network is a bidelta network. Hence, the topological equivalence of all bidelta networks implies that the cube-connected-cycles network CCC_w is topologically equivalent to the block network Block_w.

4.4.3 The Randomized Cube-Connected-Cycles Network

We prove a conditional concentration property for the randomized cube-connected-cycles network.

Lemma 4.7 *Consider the randomized cube-connected-cycles network CCC_w and fix a layer ℓ with $1 \leq \ell \leq \lg w - \lceil \lg \lg w \rceil$, an integer ζ with $0 \leq \zeta \leq \lg w - \lceil \lg \lg w \rceil - \ell + 1$, and a pair of binary strings $u_1 \in \{0,1\}^{\ell-1}$ and $u_2 \in \{0,1\}^{\lg w - \lceil \lg \lg w \rceil - \ell - \zeta + 1}$. Fix an input vector $\mathbf{x}(\ell)$ with*

$$\left| \frac{\sum_{\{u_1 u u_2 | u \in \{0,1\}^{\lceil \lg \lg w \rceil + \zeta}\}} \mathbf{x}}{2^{\lceil \lg \lg w \rceil + \zeta}} - \frac{\sum \mathbf{x}}{w} \right| \leq 2\,.$$

Consider a balancer b in layer $\ell(\mathsf{b}) = \ell + \lceil \lg \lg w \rceil + \zeta$ with input wires $i = u_1 \widehat{u} u_2$ and $i(\ell(\mathsf{b}))$, for some binary string $\widehat{u} \in \{0,1\}^{\lceil \lg \lg w \rceil + \zeta}$. Then, for any number $\delta > 0$,

$$\mathbb{P}\left[\left| x(\mathsf{b}) - \frac{\sum \mathbf{x}}{w} \right| \geq \delta \right] \leq 4 \cdot \mathsf{exp}(-(\delta - 2)^2)\,.$$

Proof: By the triangle inequality and the assumption on $\mathbf{x}(\ell)$, it follows that

$$\begin{aligned}
& \mathbb{P}\left[\left| x(\mathsf{b}) - \frac{\sum \mathbf{x}}{w} \right| \geq \delta \right] \\
\leq \quad & \mathbb{P}\left[\left| x(\mathsf{b}) - \frac{\sum_{\{u_1 u u_2 | u \in \{0,1\}^{\lceil \lg \lg w \rceil + \zeta}\}} \mathbf{x}}{2^{\lceil \lg w \lg w \rceil + \zeta}} \right| + \left| \frac{\sum_{\{u_1 u u_2 | u \in \{0,1\}^{\lceil \lg \lg w \rceil + \zeta}\}} \mathbf{x}}{2^{\lceil \lg w \lg w \rceil + \zeta}} - \frac{\sum \mathbf{x}}{w} \right| \geq \delta \right] \\
\leq \quad & \mathbb{P}\left[\left| x(\mathsf{b}) - \frac{\sum_{\{u_1 u u_2 | u \in \{0,1\}^{\lceil \lg \lg w \rceil + \zeta}\}} \mathbf{x}}{2^{\lceil \lg w \lg w \rceil + \zeta}} \right| \geq \delta - 2 \right].
\end{aligned}$$

By Lemma 4.4, the network $\mathsf{CCC}_w \setminus [\ell - 1, \lg w - \lceil \lg \lg w \rceil - \ell - \zeta + 1]$ restricted to the set of wires $\left\{u_1 u u_2 \mid u \in \{0,1\}^{\lceil \lg \lg w \rceil + \zeta}\right\}$ is a cube-connected-cycles network $\mathsf{CCC}_{2^{\lceil \lg \lg w \rceil + \zeta}}$; by definition, the input wires to balancer b are output wires of this network. Hence, by Lemma 4.3,

$$\mathbb{P}\left[\left| x(\mathsf{b}) - \frac{\sum_{\{u_1 u u_2 | u \in \{0,1\}^{\lceil \lg \lg w \rceil + \zeta}\}} \mathbf{x}}{2^{\lceil \lg w \lg w \rceil + \zeta}} \right| \geq \delta - 2 \right] \leq 4 \cdot \mathsf{exp}\left(-(\delta - 2)^2 \right).$$

It follows that

$$\mathbb{P}\left[\left| x(\mathsf{b}) - \frac{\sum \mathbf{x}}{w} \right| \geq \delta \right] \leq 4 \cdot \mathsf{exp}\left(-(\delta - 2)^2 \right).$$

as needed. ■

5 One Block

Upper and lower bounds on the smoothness of the randomized block network are presented in Sections 5.1 and 5.2, respectively.

5.1 Upper Bound

We show:

Theorem 5.1 *The randomized network* Block_w *is a* $(\lceil \lg \lg w \rceil + 3)$*-smoothing network with probability at least* $1 - \dfrac{4}{w^3}$.

Proof: Fix an input vector $\mathbf{x}$. By Lemma 4.2,

$$\begin{aligned} y_0 &= \frac{\sum \mathbf{x}}{w} + \sum_{\ell=1}^{\lg w} \frac{1}{2^{\lg w - \ell}} \sum_{\mathsf{b} \in v_{\ell,1}} \chi_{\mathsf{b}} \\ &= \frac{\sum \mathbf{x}}{w} + \sum_{\ell=1}^{\lg w - \lceil \lg \lg w \rceil} \frac{1}{2^{\lg w - \ell}} \sum_{\mathsf{b} \in v_{\ell,1}} \chi_{\mathsf{b}} + \sum_{\ell = \lg w - \lceil \lg \lg w \rceil + 1}^{\lg w} \frac{1}{2^{\lg w - \ell}} \sum_{\mathsf{b} \in v_{\ell,1}} \chi_{\mathsf{b}} \\ &= \frac{\sum \mathbf{x}}{w} + \mathsf{V}_1 + \mathsf{V}_2 \,, \end{aligned}$$

where

$$\begin{aligned} \mathsf{V}_1 &= \mathsf{V}\left(\bigcup_{1 \le \ell \le \lg w - \lg \lg w} \{ \mathsf{b} \mid \mathsf{b} \in v_{\ell,1} \} \right) \\ &= \sum_{\ell=1}^{\lg w - \lceil \lg \lg w \rceil} \frac{1}{2^{\lg w - \ell}} \sum_{\mathsf{b} \in v_{\ell,1}} \chi_{\mathsf{b}} \end{aligned}$$

and

$$\begin{aligned} \mathsf{V}_2 &= \mathsf{V}\left(\bigcup_{\lg w - \lg \lg w + 1 \le \ell \le \lg w} \{ \mathsf{b} \mid \mathsf{b} \in v_{\ell,1} \} \right) \\ &= \sum_{\ell = \lg w - \lceil \lg \lg w \rceil + 1}^{\lg w} \frac{1}{2^{\lg w - \ell}} \sum_{\mathsf{b} \in v_{\ell,1}} \chi_{\mathsf{b}} \,. \end{aligned}$$

Similarly to V_1, denote

$$\begin{aligned} \mathsf{W}_1 &= \mathsf{W}\left(\bigcup_{1 \le \ell \le \lg w - \lg \lg w} \{ \mathsf{b} \mid \mathsf{b} \in v_{\ell,1} \} \right) \\ &= \sum_{\ell=1}^{\lg w - \lceil \lg \lg w \rceil} \frac{1}{2^{\lg w - \ell}} \sum_{\mathsf{b} \in v_{\ell,1}} r_{\mathsf{b}} \,. \end{aligned}$$

For each pair of a layer ℓ with $1 \leq \ell \leq \lg w - \lceil \lg \lg w \rceil$ and a balancer $\mathsf{b} \in v_{\ell,1}$, the random variable $\frac{1}{2^{\lg w - \ell}} \cdot r_{\mathsf{b}}$ from the sum W_1 has range $\left\{ -\frac{1}{2^{\lg w - \ell + 1}}, +\frac{1}{2^{\lg w - \ell + 1}}, \right\}$. Hence, by the Hoeffding Bound (Lemma 2.4) and Claim 4.1, it follows that

$$\begin{aligned}
\mathbb{P}\left[|\mathsf{W}_1| \geq 2\right] &\leq 2 \cdot \mathsf{exp}\left(- \frac{2 \cdot 2^2}{\sum_{\ell=1}^{\lg w - \lceil \lg \lg w \rceil} \sum_{\mathsf{b} \in v_{\ell,1}} \left(\frac{1}{2^{\lg w - \ell + 1}} - \left(- \frac{1}{2^{\lg w - \ell + 1}} \right) \right)^2} \right) \\
&= 2 \cdot \mathsf{exp}\left(- \frac{2 \cdot 2^2}{\frac{2}{\lg w}} \right) \\
&\leq \frac{2}{w^4} .
\end{aligned}$$

Hence, Claim 3.8 implies that

$$\mathbb{P}\left[|\mathsf{V}_1| \geq 2\right] \leq \frac{4}{w^4} .$$

On the other hand, by the triangle inequality,

$$\begin{aligned}
|\mathsf{V}_2| &\leq \sum_{\ell = \lg w - \lceil \lg \lg w \rceil + 1}^{\lg w} \frac{1}{2^{\lg w - \ell}} \sum_{\mathsf{b} \in v_{\ell,1}} |\chi_{\mathsf{b}}| \\
&\leq \sum_{\ell = \lg w - \lceil \lg \lg w \rceil + 1}^{\lg w} \frac{1}{2^{\lg w - \ell}} \sum_{\mathsf{b} \in v_{\ell,1}} \frac{1}{2} \\
&= \sum_{\ell = \lg w - \lceil \lg \lg w \rceil + 1}^{\lg w} \frac{1}{2^{\lg w - \ell}} \cdot \frac{w}{2^{\ell}} \cdot \frac{1}{2} \\
&= \frac{1}{2} \cdot \lceil \lg \lg w \rceil .
\end{aligned}$$

We now prove a lower bound on some concentration probability for the output vector $\mathsf{Block}_w(\mathbf{x})$.

Lemma 5.2 *It holds that*

$$\mathbb{P}\left[\mathsf{Block}_w(\mathbf{x}) \textit{ is } \left(\tfrac{1}{2} \lceil \lg \lg w \rceil + 2\right)\textit{-concentrated}\right] \geq 1 - \frac{4}{w^3} .$$

Proof: By the Union Bound (Lemma 2.3),

$$
\begin{aligned}
& \mathbb{P}\left[\mathsf{Block}_w(\mathbf{x}) \text{ is } \left(\tfrac{1}{2}\lceil \lg\lg w\rceil + 2\right)\text{-concentrated}\right] \\
= \;& 1 - \mathbb{P}\left[\mathsf{Block}_w(\mathbf{x}) \text{ is not } \left(\tfrac{1}{2}\lceil \lg\lg w\rceil + 2\right)\text{-concentrated}\right] \\
= \;& 1 - \mathbb{P}\left[\bigvee_{j\in[w]} \left(\left|y_j - \frac{\sum \mathbf{x}}{w}\right| \geq \frac{1}{2}\lceil \lg\lg w\rceil + 2\right)\right] \\
\geq \;& 1 - \sum_{j\in[w]} \mathbb{P}\left[\left|y_j - \frac{\sum \mathbf{x}}{w}\right| \geq \frac{1}{2}\lceil \lg\lg w\rceil + 2\right].
\end{aligned}
$$

Since each output y_j with $j \in [w]$ is identically distributed with y_0. it follows by the Union Bound (Lemma 2.3) the triangle inequality and Lemma 2.6 that

$$
\begin{aligned}
& \mathbb{P}\left[\mathsf{Block}_w(\mathbf{x}) \text{ is } \left(\tfrac{1}{2}\lceil \lg\lg w\rceil + 2\right)\text{-concentrated}\right] \\
\geq \;& 1 - w \cdot \mathbb{P}\left[\left|y_0 - \frac{\sum \mathbf{x}}{w}\right| \geq \frac{1}{2}\lceil \lg\lg w\rceil + 2\right] \\
= \;& 1 - w \cdot \mathbb{P}\left[\left|\mathsf{V}_1 + \mathsf{V}_2\right| \geq \frac{1}{2}\lceil \lg\lg w\rceil + 2\right] \\
\geq \;& 1 - w \cdot \mathbb{P}\left[\left|\mathsf{V}_1\right| + \left|\mathsf{V}_2\right| \geq \frac{1}{2}\lceil \lg\lg w\rceil + 2\right] \\
\geq \;& 1 - w \cdot \left(\mathbb{P}\left[\left|\mathsf{V}_1\right| \geq 2\right] + \mathbb{P}\left[\left|\mathsf{V}_2\right| > \frac{1}{2}\lceil \lg\lg w\rceil\right]\right) \\
\geq \;& 1 - w \cdot \left(\frac{4}{w^4} + 0\right) \\
= \;& 1 - \frac{4}{w^3},
\end{aligned}
$$

as needed. ■

Since $2\left(\frac{1}{2}\lceil \lg\lg w\rceil + 2\right) = \lceil \lg\lg w\rceil + 4$ is an integer, it follows from Lemmas 2.2, and 5.2 that

$$
\mathbb{P}\left[\mathsf{Block}_w(\mathbf{x}) \text{ is } (\lceil \lg\lg w\rceil + 3)\text{-smooth}\right] \quad \geq \quad 1 - \frac{4}{w^3}.
$$

as needed. ■

5.2 Lower Bound

We show:

Theorem 5.3 *The randomized network* Block_w *is a* $(\lfloor \lg\lg w\rfloor - 2)$*-smoothing network with probability at most* $2 \cdot \exp\left(-\frac{4\sqrt{w}}{\lg w}\right)$.

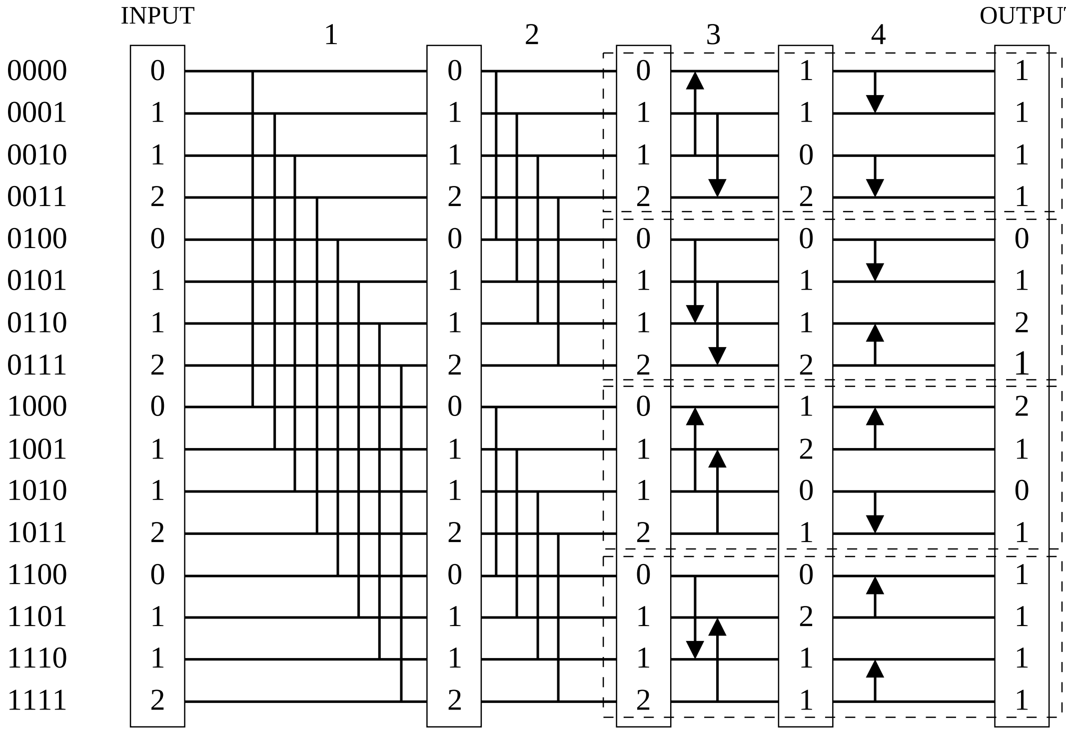

Figure 6: The network CCC_{16} with a particular orientation. The five vertical boxes contain the numbers of tokens entering on an input wire in one of the layers 1, 2, 3 and 4, or exiting on an output wire, respectively. The numbers of tokens entering on an input wire of the network are chosen as in the proof of Theorem 5.3; specifically, the number of tokens on an input wire is the number of occurences of 1 in the two least significant bits of the wire's binary representation. We shall establish that the input vector is a fixed-point for the prefix of CCC_{16} consisting of the two leftmost layers (Lemma 5.4); hence, the orientation of balancers in this prefix is not indicated.

In the proof, we shall deal with the network CCC_w. We shall construct an input vector $\mathbf{x}$ such that the probability that $\mathsf{CCC}_w(\mathbf{x})$ is $(\lfloor \lg \lg w \rfloor - 2)$-smooth is at most $2 \cdot \mathsf{exp}\left(-\frac{4\sqrt{w}}{\lg w}\right)$. Figure 6 provides an illustration for the construction used in the proof.

Proof: Construct the input vector $\mathbf{x}$ as follows. For each input wire $i = i_1 i_2 \ldots i_{\lg w}$, set

$$x_i(1) \quad := \sum_{k=\lg w - \lfloor \lg \lg w \rfloor + 2}^{\lg w} i_k \,;$$

so, x_i is the number of occurences of 1 in the $\lfloor \lg \lg w \rfloor - 1$ least significant bits of the binary representation $i_1 i_2 \ldots i_{\lg w}$. We prove:

Lemma 5.4 *For each balancer* b *in the randomized network* $\mathsf{Prefix}_{\lg w - \lfloor \lg\lg w \rfloor + 1}(\mathsf{CCC}_w)$*,*

$$x_1(\mathsf{b}) \quad = \quad x_2(\mathsf{b})\,.$$

Proof: By induction on the layer $\ell(\mathsf{b})$. For the basis case where $\ell(\mathsf{b}) = 1$, consider a balancer b in layer 1 connecting wires $i \in \{0,1\}^{\lg w}$ and $i(1)$. By construction of the input vector $\mathbf{x}$,

$$x_1(\mathsf{b}) \quad = \quad \sum_{k=\lg w - \lfloor \lg\lg w \rfloor + 2}^{\lg w} i_k$$

and

$$x_2(\mathsf{b}) \quad = \quad \sum_{k=\lg w - \lfloor \lg\lg w \rfloor + 2}^{\lg w} i(1)_k\,.$$

By construction of the network CCC_w, i and $i(1)$ differ only in bit 1; so, $i_k = i(1)_k$ for all entries $k \geq \lg w - \lfloor \lg\lg w \rfloor + 2$. Hence, $x_1(\mathsf{b}) = x_2(\mathsf{b})$, and the claim follows.

Assume inductively that the claim holds for all layers $1, \ldots, \ell - 1$, where $1 < \ell < \lg w - \lfloor \lg\lg w \rfloor + 2$. For the induction step, consider a balancer b in layer ℓ, which connects wires $i \in \{0,1\}^{\lg w}$ and $i(\ell)$. Induction hypothesis and the definition of balancer imply that for each wire $i' \in \{0,1\}^{\lg w}$,

$$x_{i'}(\ell - 1) \quad = \quad \sum_{k=\lg w - \lfloor \lg\lg w \rfloor + 2}^{\lg w} i'_k\,.$$

It follows that

$$x_1(\mathsf{b}) \quad = \quad \sum_{k=\lg w - \lfloor \lg\lg w \rfloor + 2}^{\lg w} i_k$$

and

$$x_2(\mathsf{b}) \quad = \quad \sum_{k=\lg w - \lfloor \lg\lg w \rfloor + 2}^{\lg w} i(\ell)_k\,,$$

By construction of the network, i and $i(\ell)$ differ only in bit ℓ; so, $i_k = i(\ell)_k$ for all entries $k \geq \lg w - \lfloor \lg\lg w \rfloor + 2$. Hence, $x_1(\mathsf{b}) = x_2(\mathsf{b})$, and the claim follows. ∎

Note that by Lemma 5.4, $\mathbf{x}$ is a fixed-point input of the network $\mathsf{Prefix}_{\lg w - \lfloor \lg\lg w \rfloor + 1}(\mathsf{CCC}_w)$. We now focus on the suffix network $\mathsf{Suffix}_{\lfloor \lg\lg w \rfloor - 1}(\mathsf{CCC}_w)$. By Corollary 4.5 (with $\ell_1 = \lg w - (\lceil \lg\lg w \rceil - 1)$ and $\ell_2 = 0$), the suffix $\mathsf{Prefix}_{\lfloor \lg\lg w \rfloor - 1}(\mathsf{CCC}_w)$. is the parallel cascade of $2^{\lg w - (\lceil \lg\lg w \rceil - 1)} = \frac{w}{2^{\lfloor \lg\lg w \rfloor - 1}}$ copies of the network $\mathsf{CCC}_{2^{\lg w - (\lg w - (\lfloor \lg\lg w \rfloor - 1))}} = \mathsf{CCC}_{2^{\lfloor \lg\lg w \rfloor - 1}}$.

Fix now a binary string $u \in \{0,1\}^{\lg w - \lfloor \lg\lg w \rfloor + 1}$, and consider the copy $\mathsf{CCC}_{2^{\lfloor \lg\lg w \rfloor - 1}} = \mathsf{CCC}_{2^{\lfloor \lg\lg w \rfloor - 1}}(u)$ restricted to the set of wires $\{uv \mid v \in \{0,1\}^{\lfloor \lg\lg w - 1 \rfloor}\}$. Consider the input wire $i = uv$ for some binary string $v \in \{0,1\}^{\lfloor \lg\lg w \rfloor - 1}$. By Lemma 5.4,

$$x_i \;=\; \sum_{k=\lg w - \lfloor \lg\lg w \rfloor + 2}^{\lg w} i_k \,.$$

We prove:

Lemma 5.5 *It holds that*

$$\mathbb{P}\left[y_{u1^{\lfloor \lg\lg w \rfloor - 1}} = 0\right] \;\geq\; \frac{1}{2^{(2^{\lfloor \lg\lg w \rfloor - 1} - 1)}}$$

and

$$\mathbb{P}\left[y_{u1^{\lfloor \lg\lg w \rfloor - 1}} = \lfloor \lg\lg w \rfloor - 1\right] \;\geq\; \frac{1}{2^{(2^{\lfloor \lg\lg w \rfloor - 1} - 1)}} \,.$$

Proof: Clearly,

$$\begin{aligned}\mathsf{D}_{u1^{\lfloor \lg\lg w \rfloor - 1}}[\lg w - (\lceil \lg\lg w \rceil - 1), \lg w] \;&=\; \sum_{k=0}^{\lfloor \lg\lg w \rfloor - 2} 2^k \\ &=\; 2^{\lfloor \lg\lg w \rfloor - 1} - 1 \,.\end{aligned}$$

Hence, there are $2^{2^{\lfloor \lg\lg w \rfloor - 1} - 1}$ orientations for balancers in the set $\mathsf{D}_{u1^{\lfloor \lg\lg w \rfloor - 1}}[\lg w - (\lceil \lg\lg w \rceil - 1), \lg w]$. Since these orientations occur with uniform probability, it suffices to prove that the random variable $y_{u1^{\lfloor \lg\lg w \rfloor - 1}}$ attains the values 0 and $\lfloor \lg\lg w \rfloor - 1$.

By induction on $w' = 2^{\lfloor \lg\lg w \rfloor}$. For the basis case where $w' = 4$, the claim is verified directly. Assume inductively that the random variable $y_{u1^{\lg w'}}$ in the network $\mathsf{CCC}_{w'}$ attains the values 0 and $\lg w'$. For the induction step, consider the network $\mathsf{CCC}_{2w'}$, and its output wire $0u1^{\lg w'}$. By construction of thr cube-connected-cycles network, $\mathsf{CCC}_{2w'}$ is the sequential cascade of a ladder network L_w followed by the parallel cascade of two copies of the network $\mathsf{CCC}_{w'}$; we shall focus on the top such copy.

- Assume that all balancers in layer 1 of the network $\mathsf{CCC}_{2w'}$ are oriented bottom. By construction of the input vector $\mathbf{x}$, this implies that the input to each input wire $0i'$ of $\mathsf{CCC}_{w'}$, where $i' \in \{0,1\}^{\lg(2w')-1} = \{0,1\}^{\lg w'}$, equals $\mathtt{1}(0i') = \mathtt{1}(i')$. Induction hypothesis implies that $y_{u1^{\lg w'}}$ attains the value 0.

- Assume now that all balancers in layer 1 of the network $\mathsf{CCC}_{2w'}$ are oriented top. By construction of the input vector $\mathbf{x}$, this implies that the input to each input wire $1i'$ of $\mathsf{CCC}_{w'}$, where $i' \in \{0,1\}^{\lg(2w')-1} = \{0,1\}^{\lg w'}$, equals $\mathtt{1}(1i') = 1 + \mathtt{1}(i')$. Induction hypothesis implies that $y_{u1^{\lg w'}}$ attains the value $1 + \lg w' = \lg 2w'$.

The induction is now complete. ■

We continue to prove:

Lemma 5.6 *The set* $\{y_{u1^{\lfloor \lg\lg w\rfloor -1}} \mid u \in \{0,1\}^{\lg w - \lfloor \lg\lg w\rfloor +1}\}$ *is a set of independent random variables.*

Proof: For each binary string $u \in \{0,1\}^{\lg w-(\lg\lg w-1)}$, the construction of the network CCC_w implies that the random variable $y_{u1^{\lfloor \lg\lg w\rfloor -1}}$ is determined by *(i)* the inputs to the copy $\mathsf{CCC}_{2^{\lfloor \lg\lg w\rfloor -1}}[u]$, which are fixed (by Lemma 5.4); *(ii)* the (randomly chosen) orientation of the copy $\mathsf{CCC}_{2^{\lfloor \lg\lg w\rfloor -1}}[u]$. By Corollary 4.5, the copies $\mathsf{CCC}_{2^{\lfloor \lg\lg w\rfloor -1}}[u']$ over all binary strings $u' \in \{0,1\}^{\lg w-(\lg\lg w-1)}$ are disjoint. Hence, the claim follows. ■

Lemmas 5.5 and 5.6 imply that for each $y \in \{0, \lfloor \lg\lg w\rfloor - 1\}$,

$$\begin{aligned}
&\mathbb{P}\left[\bigwedge_{u\,\in\,\{0,1\}^{\lg w\,-\,(\lfloor \lg\lg w\rfloor\,-\,1)}} \left(y_{u1^{\lfloor \lg\lg w\rfloor\,-\,1}} \neq y\right)\right] \\
=\;& \prod_{u\,\in\,\{0,1\}^{\lg w\,-\,(\lfloor \lg\lg w\rfloor\,-\,1)}} \left(1-\mathbb{P}\left[y_{u1^{\lfloor \lg\lg w\rfloor\,-\,1}} = y\right]\right) \\
\leq\;& \left(1-2^{-(2^{\lfloor \lg\lg w\rfloor\,-\,1}-1)}\right)^{2^{\lg w\,-\,(\lfloor \lg\lg w\rfloor\,-\,1)}} \\
=\;& \left(\left(1-2^{-(2^{\lfloor \lg\lg w\rfloor -1}\,-\,1)}\right)^{2^{2^{\lfloor \lg\lg w\rfloor\,-\,1}}\,-\,1}\right)^{\frac{2^{\lg w\,-\,(\lfloor \lg\lg w\rfloor\,-\,1}}{2^{2^{\lfloor \lg\lg w\rfloor\,-\,1}}\,-\,1}} \\
\leq\;& \mathsf{exp}\left(-2^{\lg w\,-\,(\lfloor \lg\lg w\rfloor\,-\,1)\,-\,2^{\lfloor \lg\lg w\rfloor\,-\,1}\,+\,1}\right) \\
\leq\;& \mathsf{exp}\left(-2^{\lg w\,-\,(\lg\lg w\,-\,1)\,-\,2^{\lg\lg w\,-\,1}\,+\,1}\right) \\
=\;& \mathsf{exp}\left(-\frac{4\sqrt{w}}{\lg w}\right).
\end{aligned}$$

Hence, by the Union Bound (Lemma 2.3),

$$\begin{aligned}
&\mathbb{P}\left[\mathbf{y}\text{ is }(\lfloor \lg\lg w\rfloor - 2)\text{-smooth}\right] \\
=\;& \mathbb{P}\left[\left(\bigwedge_{u\in\{0,1\}^{\lg w-(\lfloor \lg\lg w\rfloor+1}} (y_{u1^{\lfloor \lg\lg w\rfloor -1}} \neq 0)\right)\bigvee\left(\bigwedge_{u\in\{0,1\}^{\lg w-(\lfloor \lg\lg w\rfloor+1}} (y_{u1^{\lfloor \lg\lg w\rfloor -1}} \neq \lfloor \lg\lg w\rfloor - 1)\right)\right] \\
\leq\;& \mathbb{P}\left[\bigwedge_{u\in\{0,1\}^{\lg w-(\lfloor \lg\lg w\rfloor-1)}} (y_{u1^{\lfloor \lg\lg w\rfloor -1}} \neq 0)\right] + \mathbb{P}\left[\bigwedge_{u\in\{0,1\}^{\lg w-\lfloor \lg\lg w\rfloor+1}} (y_{u1^{\lfloor \lg\lg w\rfloor -1}} \neq \lfloor \lg\lg w\rfloor - 1)\right] \\
\leq\;& 2\cdot\mathsf{exp}\left(-\frac{4\sqrt{w}}{\lg w}\right),
\end{aligned}$$

as needed. ■

6 Two Blocks

Henceforth, we shall assume that $w \geq 2^{2^{18}}$ since otherwise our main result for two blocks (Theorem 6.7) follows from Theorem 5.1. Since the block network Block_w and the cube-connected-cycles network CCC_w are topologically equivalent, so are Block^2_w and CCC^2_w. Hence, we shall deal throughout with the cascade of two cube-connected-cycles networks.

The rest of this section is organized as follows. Section 6.1 establishes the *Concentration-to-Average Lemma*. Section 6.2 presents and analyzes a partitioning of the second cascaded cube-connected-cycles network into groups. Section 6.3 establishes some concentration properties for each of the partitioned groups. Our main result for two blocks is established in Section 6.4.

6.1 The Concentration-to-Average Lemma

We start with a significant definition:

Definition 6.1 (Concentration-to-Average) *Fix an input vector* $\mathbf{x}$ *and a layer* ℓ *with* $\lg w + 1 \leq \ell \leq 2\lg w$ *in the network* CCC^2_w. *Denote as* $\mathcal{E}(\mathbf{x}, \ell)$ *the event that for all integers* ζ *with* $0 \leq \zeta \leq 2\lg w - \lfloor \lg\lg w \rfloor - \ell + 1$, *and for all pairs of binary strings* $u_1 \in \{0,1\}^{\ell - \lg w - 1}$ *and* $u_2 \in \{0,1\}^{2\lg w - \ell - \lceil \lg\lg w \rceil - \zeta + 1}$,

$$\left| \frac{\sum_{\{u_1 u u_2 \mid u \in \{0,1\}^{\lceil \lg\lg w \rceil + \zeta}\}} \mathbf{x}(\ell)}{2^{\lceil \lg\lg w \rceil + \zeta}} - \frac{\sum \mathbf{x}}{w} \right| \quad \leq \quad 2\,.$$

Recall that by Lemma 4.4, for each integer ζ with $0 \leq \zeta \leq 2\lg w - \lceil \lg\lg w \rceil - \ell + 1$, and for each pair of binary strings $u_1 \in \{0,1\}^{\ell - \lg w - 1}$ and $u_2 \in \{0,1\}^{2\lg w - \ell - \lceil \lg\lg w \rceil - \zeta + 1}$, the network $\mathsf{CCC}_w \setminus [\ell - 1, \lg w - \lceil \lg\lg w \rceil - \ell - \zeta + 1]$ restricted to the set of wires $\left\{ u_1 u u_2 \mid u \in \{0,1\}^{\lceil \lg\lg w \rceil + \zeta} \right\}$ is a cube-connected-cycles network $\mathsf{CCC}_{2^{\lceil \lg\lg w \rceil + \zeta}}$. Hence, the event $\mathcal{E}(\mathbf{x}, \ell)$ refers to each separate restriction of the inputs (to CCC^2_w) on the input wires of such a cube-connected-cycles network; it asserts that each such cube-connected-cycles network receives an average number of tokens (over its input wires) which is within 2 of the average $\frac{\sum \mathbf{x}}{w}$. We continue to establish that the event $\mathcal{E}(\mathbf{x}, \ell)$ occurs with high probability. We prove:

Lemma 6.1 (Concentration-to-Average Lemma) *Consider the randomized cube-connected-cycles network* CCC^2_w. *Fix an input vector* $\mathbf{x}$ *and a layer* ℓ *in the network* CCC^2_w, *with* $\lg w + 1 \leq \ell \leq 2\lg w$. *Then,*

$$\mathbb{P}\left[\mathcal{E}(\mathbf{x}, \ell)\right] \quad \geq \quad 1 - \frac{4}{w^3}\,.$$

Proof: Fix an integer ζ with with $0 \leq \zeta \leq 2\lg w - \lceil \lg\lg w \rceil - \ell + 1$, and a pair of binary strings $u_1 \in \{0,1\}^{\ell - \lg w - 1}$ and $u_2 \in \{0,1\}^{2\lg w - \ell - \lceil \lg\lg w \rceil - \zeta + 1}$. Since $\ell \geq \lg w + 1$, it follows that $\zeta \leq \lg w - \lceil \lg\lg w \rceil$.

Fix a wire $i = u_1 u u_2 \in \{0,1\}^{\lg w}$, for some (fixed) binary string $u \in \{0,1\}^{\lceil \lg\lg w \rceil + \zeta}$. Consider layer $\ell' = \ell - \lg w + \lceil \lg\lg w \rceil + \zeta - 1$. By assumption on ζ, it follows that $\ell' \leq \ell - \lg w + \lceil \lg\lg w \rceil + 2\lg w - \lfloor \lg w \rfloor - \ell + 1 - 1 = \lg w$, so that ℓ' is a layer in the first cascaded CCC_w network. By Lemma 4.6, this implies that

$$\mathsf{ID}_i(\ell') \quad = \quad \left\{ \widehat{u}_1 u \widehat{u}_2 \mid \widehat{u}_1 \in \{0,1\}^{\ell - 1 - \lg w} \text{ and } \widehat{u}_2 \in \{0,1\}^{\lg w - \ell' + 1} \right\}.$$

Note that that the network CCC_w^2 is complete; hence, we apply Lemma 3.7 on the (complete) network $\mathsf{Pref}_{\ell-1}\mathsf{CCC}_w^2$ to obtain that

$$\begin{aligned}
x_{u_1 u u_2}(\ell) &= y_i(\ell - 1) \\
&= \frac{1}{2^{(\ell-1)-\ell'}} \sum_{i' \in \mathsf{ID}_i(\ell')} y_{i'}(\ell') + \sum_{k=\ell'+1}^{\ell-1} \frac{1}{2^{(\ell-1)-k}} \sum_{\mathsf{b} \in \mathsf{D}_i[k]} \chi_{\mathsf{b}} \\
&= \frac{1}{2^{\lg w - \lceil \lg\lg w \rceil - \zeta}} \sum_{i' \in \mathsf{ID}_i(\ell')} x_{i'}(\ell' + 1) + \sum_{k=\ell'+1}^{\ell-1} \frac{1}{2^{(\ell-1)-k}} \sum_{\mathsf{b} \in \mathsf{D}_i[k]} \chi_{\mathsf{b}} \\
&= \frac{1}{2^{\lg w - \lceil \lg\lg w \rceil - \zeta}} \sum_{\substack{\widehat{u}_1 \in \{0,1\}^{\ell-1-\lg w} \\ \widehat{u}_2 \in \{0,1\}^{\lg w - \ell' + 1}}} x_{\widehat{u}_1 u \widehat{u}_2}(\ell' + 1) + \sum_{k=\ell'+1}^{\ell-1} \frac{1}{2^{(\ell-1)-k}} \sum_{\mathsf{b} \in \mathsf{D}_{u_1 u u_2}[k]} \chi_{\mathsf{b}}\,.
\end{aligned}$$

Summing over all binary strings $u \in \{0,1\}^{\lceil \lg\lg w \rceil + \zeta}$, we obtain that

$$\begin{aligned}
& \sum_{\{u_1 u u_2 \mid u \in \{0,1\}^{\lceil \lg\lg w \rceil + \zeta}\}} \mathbf{x}(\ell) \\
= \quad & \sum_{u \in \{0,1\}^{\lceil \lg\lg w \rceil + \zeta}} \left(\frac{1}{2^{\lg w - \lceil \lg\lg w \rceil - \zeta}} \sum_{\substack{\widehat{u}_1 \in \{0,1\}^{\ell-1-\lg w} \\ \widehat{u}_2 \in \{0,1\}^{\lg w - \ell' + 1}}} x_{\widehat{u}_1 u \widehat{u}_2}(\ell' + 1) + \sum_{k=\ell'+1}^{\ell-1} \frac{1}{2^{(\ell-1)-k}} \sum_{\mathsf{b} \in \mathsf{D}_{u_1 u u_2}[k]} \chi_{\mathsf{b}} \right) \\
= \quad & \frac{1}{2^{\lg w - \lceil \lg\lg w \rceil - \zeta}} \sum_{u \in \{0,1\}^{\lceil \lg\lg w \rceil + \zeta}} \sum_{\substack{\widehat{u}_1 \in \{0,1\}^{\ell-1-\lg w} \\ \widehat{u}_2 \in \{0,1\}^{\lg w - \ell' + 1}}} x_{\widehat{u}_1 u \widehat{u}_2}(\ell' + 1) \\
& + \sum_{u \in \{0,1\}^{\lceil \lg\lg w \rceil + \zeta}} \sum_{k=\ell'+1}^{\ell-1} \frac{1}{2^{(\ell-1)-k}} \sum_{\mathsf{b} \in \mathsf{D}_{u_1 u u_2}[k]} \chi_{\mathsf{b}} \\
= \quad & \frac{1}{2^{\lg w - \lceil \lg\lg w \rceil - \zeta}} \sum \mathbf{x}(\ell' + 1) + \sum_{u \in \{0,1\}^{\lceil \lg\lg w \rceil + \zeta}} \sum_{k=\ell'+1}^{\ell-1} \frac{1}{2^{(\ell-1)-k}} \sum_{\mathsf{b} \in \mathsf{D}_{u_1 u u_2}[k]} \chi_{\mathsf{b}} \\
= \quad & \frac{1}{2^{\lg w - \lceil \lg\lg w \rceil - \zeta}} \sum \mathbf{x} + \sum_{u \in \{0,1\}^{\lceil \lg\lg w \rceil + \zeta}} \sum_{k=\ell'+1}^{\ell-1} \frac{1}{2^{(\ell-1)-k}} \sum_{\mathsf{b} \in \mathsf{D}_{u_1 u u_2}[k]} \chi_{\mathsf{b}}\,.
\end{aligned}$$

Hence,

$$
\begin{aligned}
& \frac{\sum_{\{u_1 u u_2 \mid u \in \{0,1\}^{\lceil \lg\lg w\rceil+\zeta}\}} \mathbf{x}(\ell)}{2^{\lceil \lg\lg w\rceil+\zeta}} \\
= \; & \frac{1}{w} \cdot \sum \mathbf{x} + \frac{1}{2^{\lceil \lg\lg w\rceil+\zeta}} \cdot \sum_{u\in\{0,1\}^{\lceil \lg\lg w\rceil+\zeta}} \sum_{k=\ell'+1}^{\ell-1} \frac{1}{2^{(\ell-1)-k}} \sum_{\mathsf{b}\in \mathsf{D}_{u_1 u u_2}[k]} \chi_{\mathsf{b}} \\
= \; & \frac{\sum \mathbf{x}}{w} + \mathsf{V}(u_1, u_2)\,,
\end{aligned}
$$

where

$$
\begin{aligned}
\mathsf{V}(u_1,u_2) &= \mathsf{V}\left(\bigcup_{u\in\{0,1\}^{\lceil \lg\lg w\rceil+\zeta}} \bigcup_{k=\ell'+1}^{\ell-1} \bigcup_{\mathsf{b}\in \mathsf{D}_{u_1 u u_2}[k]} \{\mathsf{b}\} \right) \\
&= \frac{1}{2^{\lceil \lg\lg w\rceil+\zeta}} \sum_{u\in\{0,1\}^{\lceil \lg\lg w\rceil+\zeta}} \sum_{k=\ell'+1}^{\ell-1} \frac{1}{2^{(\ell-1)-k}} \sum_{\mathsf{b}\in \mathsf{D}_{u_1 u u_2}[k]} \chi_{\mathsf{b}}\,;
\end{aligned}
$$

Similarly to $\mathsf{V}(u_1, u_2)$, define the random variable

$$
\begin{aligned}
\mathsf{W}(u_1,u_2) &= \mathsf{W}\left(\bigcup_{u\in\{0,1\}^{\lceil \lg\lg w\rceil+\zeta}} \bigcup_{k=\ell'+1}^{\ell-1} \bigcup_{\mathsf{b}\in \mathsf{D}_{u_1 u u_2}[k]} \{\mathsf{b}\} \right) \\
&= \frac{1}{2^{\lceil \lg\lg w\rceil+\zeta}} \sum_{u\in\{0,1\}^{\lceil \lg\lg w\rceil+\zeta}} \sum_{k=\ell'+1}^{\ell-1} \sum_{\mathsf{b}\in \mathsf{D}_{u_1 u u_2}[k]} \frac{1}{2^{(\ell-1)-k}}\, r_{\mathsf{b}}\,.
\end{aligned}
$$

Note that by linearity of expectation, $\mathbb{E}\left[\mathsf{W}(u_1, u_2)\right] = 0$. For each triple of a layer k with $\ell' + 1 \leq k \leq \ell - 1$, a binary string $u \in \{0,1\}^{\lceil \lg\lg w\rceil+\zeta}$ and a balancer $\mathsf{b} \in \mathsf{D}_{u_1 u u_2}[k]$, define the random variable

$$
v_{\mathsf{b}} = \frac{1}{2^{\lceil \lg\lg w\rceil+\zeta}} \, \frac{1}{2^{\ell-1-k}}\, r_{\mathsf{b}}
$$

from the sum $\mathsf{W}(u_1, u_2)$; clearly, v_{b} has range

$$
\begin{aligned}
\mathsf{Range}(v_{\mathsf{b}}) &= \frac{1}{2^{\lceil \lg\lg w\rceil+\zeta}} \, \frac{1}{2^{\ell-1-k}} \left\{ -\frac{1}{2}, +\frac{1}{2} \right\} \\
&= \left\{ -\frac{1}{2^{\lceil \lg\lg w\rceil+\zeta+\ell-k}}, +\frac{1}{2^{\lceil \lg\lg w\rceil+\zeta+\ell-k}} \right\},
\end{aligned}
$$

so that

$$
|\mathsf{Range}(v_{\mathsf{b}})| = \frac{1}{2^{\lceil \lg\lg w\rceil+\zeta+\ell-k-1}}\,.
$$

We continue with an elementary technical claim.

Claim 6.2 *It holds that*

$$\sum_{u\in\{0,1\}^{\lceil\lg\lg w\rceil+\zeta}} \sum_{k=\ell'+1}^{\ell-1} \sum_{\mathsf{b}\in \mathsf{D}_{u_1uu_2}[k]} |\mathsf{Range}(v_{\mathsf{b}})|^2 \;\leq\; \frac{2}{\lg w}\,.$$

Proof: Clearly,

$$\begin{aligned}
\sum_{u\in\{0,1\}^{\lceil\lg\lg w\rceil+\zeta}} \sum_{k=\ell'+1}^{\ell-1} \sum_{\mathsf{b}\in \mathsf{D}_{u_1uu_2}[k]} |\mathsf{Range}(v_{\mathsf{b}})|^2 &= \sum_{u\in\{0,1\}^{\lceil\lg\lg w\rceil+\zeta}} \sum_{k=\ell'+1}^{\ell-1} \sum_{\mathsf{b}\in \mathsf{D}_{u_1uu_2}[k]} \frac{1}{2^{2(\lceil\lg\lg w\rceil+\zeta+\ell-k-1)}}\\
&= \sum_{u\in\{0,1\}^{\lceil\lg\lg w\rceil+\zeta}} \sum_{k=\ell'+1}^{\ell-1} \frac{1}{2^{2(\lceil\lg\lg w\rceil+\zeta+\ell-k-1)}}\; |\mathsf{D}_{u_1uu_2}[k]|\,.
\end{aligned}$$

Recall that the network CCC_w^2 is full; hence, for each pair of a binary string $u\in\{0,1\}^{\lceil\lg\lg w\rceil+\zeta}$ and an integer k with $\ell'+1\leq k\leq \ell-1$, $|\mathsf{D}_{u_1uu_2}[k]| = 2^{\ell-k-1}$. This implies that

$$\begin{aligned}
\sum_{u\in\{0,1\}^{\lceil\lg\lg w\rceil+\zeta}} \sum_{k=\ell'+1}^{\ell-1} \sum_{\mathsf{b}\in \mathsf{D}_{u_1uu_2}[k]} |\mathsf{Range}(v_{\mathsf{b}})|^2 &= \sum_{u\in\{0,1\}^{\lceil\lg\lg w\rceil+\zeta}} \sum_{k=\ell'+1}^{\ell-1} \frac{1}{2^{2(\lceil\lg\lg w\rceil+\zeta+\ell-k-1)}}\cdot 2^{\ell-k-1}\\
&= \frac{1}{2^{2(\lceil\lg\lg w\rceil+\zeta)+(\ell-1)}}\cdot \sum_{u\in\{0,1\}^{\lceil\lg\lg w\rceil+\zeta}} \sum_{k=\ell'+1}^{\ell-1} \frac{1}{2^{-k}}\\
&\leq \frac{1}{2^{2(\lceil\lg\lg w\rceil+\zeta)+(\ell-1)}}\cdot 2^{\lceil\lg\lg w\rceil+\zeta}\,2^{\ell}\\
&= \frac{2}{2^{\lceil\lg\lg w\rceil+\zeta}}\\
&\leq \frac{2}{\lg w}\,,
\end{aligned}$$

as needed. ■

By Lemma 3.8, the Hoeffding Bound (Lemma 2.4) and Claim 6.2, it follows that

$$\begin{aligned}
&\mathbb{P}\left[\left|\frac{\sum_{u_1uu_2|u\in\{0,1\}^{\lceil\lg\lg w\rceil+\zeta}}\mathbf{x}(\ell)}{2^{\lceil\lg\lg w\rceil+\zeta}} - \frac{\mathbf{x}}{w}\right| \;\geq\; 2\right]\\
=\;& \mathbb{P}\left[|\mathsf{V}(u_1,u_2)| \;\geq\; 2\right]\\
\leq\;& 2\cdot\mathbb{P}\left[|\mathsf{W}(u_1,u_2)| \;\geq\; 2\right]\\
=\;& 2\cdot\mathbb{P}\left[\left|\sum_{u\in\{0,1\}^{\lceil\lg\lg w\rceil+\zeta}} \sum_{k=\ell'+1}^{\ell-1} \sum_{\mathsf{b}\in \mathsf{D}_{u_1uu_2}[k]} v_{\mathsf{b}}\right| \;\geq\; 2\right]\\
\leq\;& 2\cdot 2\,\mathsf{exp}\left(-\frac{2\cdot 2^2}{\sum_{u\in\{0,1\}^{\lceil\lg\lg w\rceil+\zeta}}\sum_{k=\ell'+1}^{\ell-1}\sum_{\mathsf{b}\in\mathsf{D}_{u_1uu_2}}|\mathsf{Range}(v_{\mathsf{b}})|^2}\right)\\
\leq\;& 4\cdot\mathsf{exp}\left(-\frac{8}{\frac{2}{\lg w}}\right)\\
=\;& 4\cdot\mathsf{exp}\left(-4\lg w\right)\\
=\;& \frac{4}{w^4}\,.
\end{aligned}$$

Hence, by the Union Bound (Lemma 2.3),

$$
\begin{aligned}
& \mathbb{P}\left[\mathcal{E}(\mathbf{x},\ell)\right] \\
= \; & 1 - \mathbb{P}\left[\neg\mathcal{E}(\mathbf{x},\ell)\right] \\
= \; & 1 - \mathbb{P}\left[\bigvee_{\zeta=0}^{\lg w - \lceil \lg\lg w\rceil} \bigvee_{\substack{u_1 \in \{0,1\}^{\ell - \lg w - 1} \\ u_2 \in \{0,1\}^{2\lg w - \ell - \lceil \lg\lg w\rceil - \zeta + 1}}} \left| \frac{\sum_{\{u_1 u u_2 \mid u \in \{0,1\}^{\lceil \lg\lg w\rceil + \zeta}\}} \mathbf{x}(\ell)}{2^{\lceil \lg\lg w\rceil + \zeta}} - \frac{\sum \mathbf{x}}{w} \right| > 2 \right] \\
\geq \; & 1 - \sum_{\zeta=0}^{\lg w - \lceil \lg\lg w\rceil} \sum_{\substack{u_1 \in \{0,1\}^{\ell - \lg w - 1} \\ u_2 \in \{0,1\}^{2\lg w - \ell - \lceil \lg\lg w\rceil - \zeta + 1}}} \mathbb{P}\left[\left| \frac{\sum_{\{u_1 u u_2 \mid u \in \{0,1\}^{\lceil \lg\lg w\rceil + \zeta}\}} \mathbf{x}(\ell)}{2^{\lceil \lg\lg w\rceil + \zeta}} - \frac{\sum \mathbf{x}}{w} \right| > 2 \right] \\
\geq \; & 1 - \sum_{\zeta=0}^{\lg w - \lceil \lg\lg w\rceil} \sum_{\substack{u_1 \in \{0,1\}^{\ell - \lg w - 1} \\ u_2 \in \{0,1\}^{2\lg w - \ell - \lceil \lg\lg w\rceil - \zeta + 1}}} \frac{4}{w^4} \\
\geq \; & 1 - \lg w \cdot 2^{\lg w - \lceil \lg\lg w\rceil} \cdot \frac{4}{w^4} \\
= \; & 1 - \frac{4}{w^3},
\end{aligned}
$$

as needed. ■

6.2 The Partition

For each index ρ with $1 \leq \rho \leq \left\lceil \frac{\lg\lg w}{2} \right\rceil - 6$, we consider a group $\mathcal{L}_\rho$ of $\left\lceil \frac{4\lg w}{\left(\left\lceil \frac{\lg\lg w}{2}\right\rceil - 1 - \rho\right)^2} \right\rceil +$ $\lceil \lg\lg w\rceil$ consecutive layers in the second cascaded cube-connected-cycles network CCC_w. These groups are defined inductively as follows:

- For the basis case, $\mathcal{L}_1$ consists of layers $1, \ldots, \left\lceil \frac{4\lg w}{\left(\left\lceil \frac{\lg\lg w}{2}\right\rceil - 2\right)^2} \right\rceil + \lceil \lg\lg w\rceil$.
- Assume inductively that we have defined group $\mathcal{L}_{\rho-1}$, where $2 \leq \rho \leq \left\lceil \frac{\lg\lg w}{2} \right\rceil - 6$.
- For the induction step, group $\mathcal{L}_\rho$ consists of the $\left\lceil \frac{4\lg w}{\left(\left\lceil \frac{\lg\lg w}{2}\right\rceil - 1 - \rho\right)^2} \right\rceil + \lceil \lg\lg w\rceil$ layers which immediately follow group $\mathcal{L}_{\rho-1}$.

For each group $\mathcal{L}_\rho$ with $1 \leq \rho \leq \left\lceil \frac{\lg\lg w}{2} \right\rceil - 6$, denote as ℓ_ρ the leftmost layer in group $\mathcal{L}_\rho$; so, $\ell_1 = 1$, and for each group $\mathcal{L}_\rho$, $\ell_\rho < \lg w - \lceil \lg\lg w\rceil$. Also, denote as $\mathbf{x}(\rho)$ and $\mathbf{y}(\rho)$ the input and output vectors for the group $\mathcal{L}_\rho$, respectively. We calculate:

Lemma 6.3 *The total number of layers in the groups* $\mathcal{L}_\rho$ *with* $1 \le \rho \le \left\lceil \frac{\lg\lg w}{2} \right\rceil - 6$ *is less than* $\lg w$.

We shall employ the implication $(\lg\lg w)^2 \le \frac{1}{10} \cdot \lg w$ of the assumption that $w \ge 2^{2^{18}}$.

Proof: Clearly,

$$
\begin{aligned}
& \sum_{\rho=1}^{\left\lceil \frac{\lg\lg w}{2} \right\rceil - 6} \left(\left\lceil \frac{4\lg w}{\left(\left\lceil \frac{\lg\lg w}{2} \right\rceil - 1 - \rho\right)^2} \right\rceil + \lceil \lg\lg w \rceil \right) \\
\le\ & \sum_{\rho=1}^{\left\lceil \frac{\lg\lg w}{2} \right\rceil - 6} \left\lceil \frac{4\lg w}{\left(\left\lceil \frac{\lg\lg w}{2} \right\rceil - 1 - \rho\right)^2} \right\rceil + \left(\left\lceil \frac{\lg\lg w}{2} \right\rceil - 6 \right) \cdot \lceil \lg\lg w \rceil \\
\le\ & \sum_{\rho=1}^{\left\lceil \frac{\lg\lg w}{2} \right\rceil - 6} \left(\frac{4\ \lg w}{\left(\left\lceil \frac{\lg\lg w}{2} \right\rceil - 1 - \rho\right)^2} + 1 \right)^2 + \left(\left\lceil \frac{\lg\lg w}{2} \right\rceil - 6 \right) \cdot \lceil \lg\lg w \rceil \\
=\ & \sum_{\rho=1}^{\left\lceil \frac{\lg\lg w}{2} \right\rceil - 6} \frac{4\ \lg w}{\left(\left\lceil \frac{\lg\lg w}{2} \right\rceil - 1 - \rho\right)^2} + \left(\left\lceil \frac{\lg\lg w}{2} \right\rceil - 6 \right) + \left(\left\lceil \frac{\lg\lg w}{2} \right\rceil - 6 \right) \cdot \lceil \lg\lg w \rceil \\
\le\ & \sum_{\rho=1}^{\left\lceil \frac{\lg\lg w}{2} \right\rceil - 6} \frac{4\ \lg w}{\left(\left\lceil \frac{\lg\lg w}{2} \right\rceil - 1 - \rho\right)^2} + \left(\frac{\lg\lg w}{2} - 5 \right) + \left(\frac{\lg\lg w}{2} - 5 \right) \cdot (\lg\lg w + 1) \\
=\ & \sum_{\rho=1}^{\left\lceil \frac{\lg\lg w}{2} \right\rceil - 6} \frac{4\ \lg w}{\left(\left\lceil \frac{\lg\lg w}{2} \right\rceil - 1 - \rho\right)^2} + \frac{1}{2}(\lg\lg w)^2 - 4\lg\lg w - 10 \\
\le\ & 4\ \lg w \cdot \left(\sum_{k=5}^{\infty} \frac{1}{k^2} \right) + \frac{1}{2}\,(\lg\lg w)^2 \\
\le\ & 4\ \lg w \cdot \left(\sum_{k=1}^{\infty} \frac{1}{k^2} - \sum_{k=1}^{4} \frac{1}{k^2} \right) + \frac{1}{2}\,(\lg\lg w)^2 \\
=\ & 4\ \lg w \cdot \left(\frac{\pi^2}{6} - \frac{205}{144} \right) + \frac{1}{2}\,(\lg\lg w)^2 \\
\le\ & \frac{9}{10}\ \lg w + \frac{1}{2}\,(\lg\lg w)^2 \\
\le\ & \frac{9}{10}\ \lg w + \frac{1}{20}\ \lg w \\
<\ & \lg w\,,
\end{aligned}
$$

as needed. ■

Lemma 6.3 implies that for each group $\mathcal{L}_\rho$ with $1 \leq \rho \leq \left\lceil \frac{\lg\lg w}{2} \right\rceil - 6$, $|\mathsf{P}[\mathcal{L}_\rho]| < w \cdot 2^{\lg w} = w^2$.

6.3 Concentration Properties of Groups

We now prove some concentration property for each group from the partition in Section 6.2.

Lemma 6.4 *Consider group $\mathcal{L}_\rho$, where $1 \leq \rho \leq \left\lceil \frac{\lg\lg w}{2} \right\rceil - 6$, with a random input vector $\mathbf{x}$. Fix a path $\beta \in \mathsf{P}[\mathcal{L}_\rho]$. Then,*

$$\mathbb{P}\left[\bigwedge_{\mathsf{b}\in\beta}\left(\left|\widehat{x}(\mathsf{b}) - \frac{\sum \mathbf{x}}{w}\right| > \left\lceil \frac{\lg\lg w}{2} \right\rceil + 1 - \rho\right) \Big| \mathcal{E}\left(\mathbf{x}, \ell_\rho\right)\right] \quad \leq \quad \frac{4}{w^3}\,.$$

For the sake of shortening the notation, set $\gamma(\rho) := \left\lceil \frac{\lg\lg w}{2} \right\rceil + 1 - \rho$. Since $\rho \leq \left\lceil \frac{\lg\lg w}{2} \right\rceil - 6$, it follows that $\gamma(\rho) \geq 7$.

Proof: Since CCC_w is a bidelta network, Claim 3.3 implies that

$$\begin{aligned}
\mathbb{P}\left[\bigwedge_{\mathsf{b}\in\beta}\left(\left|\widehat{x}(\mathsf{b}) - \frac{\sum \mathbf{x}}{w}\right| > \gamma(\rho)\right) \Big| \mathcal{E}\left(\mathbf{x}, \ell_\rho\right)\right] \quad &\leq \quad \mathbb{P}\left[\bigwedge_{\mathsf{b}\in\beta[\lceil\lg\lg w\rceil, \mathsf{d}(\mathcal{L}_\rho)]}\left(\left|\widehat{x}(\mathsf{b}) - \frac{\sum \mathbf{x}}{w}\right| > \gamma(\rho)\right) \Big| \mathcal{E}\left(\mathbf{x}, \ell_\rho\right)\right] \\
&= \quad \prod_{\mathsf{b}\in\beta[\lceil\lg\lg w\rceil, \mathsf{d}(\mathcal{L}_\rho)]} \mathbb{P}\left[\left(\left|\widehat{x}(\mathsf{b}) - \frac{\sum \mathbf{x}}{w}\right| > \gamma(\rho)\right) \Big| \mathcal{E}\left(\mathbf{x}, \ell_\rho\right)\right]
\end{aligned}$$

Since $1 \leq \ell_\rho \leq \lg w - \lceil \lg\lg w \rceil$, Lemma 4.7 applies to yield that for each balancer $\mathsf{b} \in \beta[\lceil\lg\lg w\rceil, \mathsf{d}(\mathcal{L}_\rho)]$

$$\mathbb{P}\left[\left(\left|\widehat{x}(\mathsf{b}) - \frac{\sum \mathbf{x}}{w}\right| > \gamma(\rho)\right) \Big| \mathcal{E}\left(\mathbf{x}, \ell_\rho\right)\right] \quad \leq \quad 4 \cdot \exp\left(-(\gamma(\rho) - 2)^2\right)\,.$$

It follows that

$$
\begin{aligned}
& \mathbb{P}\left[\bigwedge_{\mathsf{b}\in\beta}\left(\left|\widehat{x}(\mathsf{b})-\frac{\sum\mathbf{x}}{w}\right|>\gamma(\rho)\right)\Big|\mathcal{E}\left(\mathbf{x},\ell_\rho\right)\right]\\
\leq\ & \left(4\cdot\mathsf{exp}\left(-(\gamma(\rho)-2)^2\right)\ .\right)^{|\beta[\lceil\lg\lg w\rceil,\,\mathsf{d}(\mathcal{L}_\rho)]|}\\
=\ & \left(\mathsf{exp}(\ln 4)\cdot\mathsf{exp}\left(-(\gamma(\rho)-2)^2\right)\ .\right)^{|\beta[\lceil\lg\lg w\rceil,\mathsf{d}(\mathcal{L}_\rho)]|}\\
=\ & \mathsf{exp}\left(|\beta[\lceil\lg\lg w\rceil,\mathsf{d}(\mathcal{L}_\rho)]|\cdot\left(\ln 4-(\gamma(\rho)-2)^2\right)\right)\\
=\ & \mathsf{exp}\left((\mathsf{d}(\mathcal{L}_\rho)-\lceil\lg\lg w\rceil)\cdot\left(\ln 4-(\gamma(\rho)-2)^2\right)\right)\\
=\ & \mathsf{exp}\left(\ln 4\cdot\left\lceil\frac{4\lg w}{(\gamma(\rho)-2)^2}\right\rceil\right)\cdot\mathsf{exp}\left(-\left\lceil\frac{4\lg w}{(\gamma(\rho)-2)^2}\right\rceil\cdot(\gamma(\rho)-2)^2\right)\\
\leq\ & \mathsf{exp}\left(\ln 4\cdot\left\lceil\frac{4\lg w}{(7-2)^2}\right\rceil\right)\cdot\mathsf{exp}\left(-4\lg w\right)\\
\leq\ & \mathsf{exp}\left(\ln 4\cdot\left\lceil\frac{4\lg w}{25}\right\rceil\right)\cdot\mathsf{exp}\left(-4\lg w\right)\\
\leq\ & \mathsf{exp}\left(\frac{4\ln 4}{25}\ \lg w+\ln 4-\lg w\right)\\
\leq\ & \mathsf{exp}\left(-3\lg w\right)\\
\leq\ & \frac{4}{w^3}\,,
\end{aligned}
$$

as needed. ■

We continue to prove:

Lemma 6.5 *Consider group $\mathcal{L}_\rho$, where $1\leq\rho\leq\left\lceil\frac{\lg\lg w}{2}\right\rceil-6$, with a (random) input vector $\mathbf{x}$. Then,*

$$
\mathbb{P}\left[\mathcal{L}_\rho(\mathbf{x})\ \textit{is not}\ \left(\left\lceil\frac{\lg\lg w}{2}\right\rceil+1-\rho\right)\textit{-concentrating on input}\ \mathbf{x}\ \right]\ <\ \frac{8}{w}\,.
$$

Proof: By Claim 3.5 (with $\mathcal{E}(\mathbf{x},\ell_\rho)$ for $\mathcal{E}$), Lemma 6.1, the Union Bound (Lemma 2.3) and

Lemma 6.4, it follows that

$$\begin{aligned}
& \mathbb{P}\left[\mathcal{L}_\rho(\mathbf{x}) \text{ is not } \left(\left\lceil\frac{\lg\lg w}{2}\right\rceil + 1 - \rho\right)\text{-concentrating on input } \mathbf{x}\ \right] \\
\leq\ & \mathbb{P}\left[\mathcal{L}_\rho(\mathbf{x}) \text{ is not } \left(\left\lceil\frac{\lg\lg w}{2}\right\rceil + 1 - \rho\right)\text{-concentrating on input } \mathbf{x}\ \middle|\mathcal{E}(\mathbf{x}, \ell_\rho)\right] + \mathbb{P}\left[\neg\mathcal{E}(\mathbf{x}, \ell_\rho)\right] \\
\leq\ & \mathbb{P}\left[\mathcal{L}_\rho(\mathbf{x}) \text{ is not } \left(\left\lceil\frac{\lg\lg w}{2}\right\rceil + 1 - \rho\right)\text{-concentrating on input } \mathbf{x}\ \middle|\mathcal{E}(\mathbf{x}, \ell_\rho)\right] + \frac{4}{w^3} \\
=\ & \mathbb{P}\left[\bigvee_{\beta\in\mathsf{P}[\mathcal{L}_\rho]} \bigwedge_{\mathsf{b}\in\beta}\left(\left|\widehat{x}(\mathsf{b}) - \frac{\sum\mathbf{x}}{w}\right| > \left\lceil\frac{\lg\lg w}{2}\right\rceil + 1 - \rho\right)\middle|\, \mathcal{E}\left(\mathbf{x}, \ell_\rho\right)\right] + \frac{4}{w^3} \\
\leq\ & |\mathsf{P}[\mathcal{L}_\rho]| \cdot \mathbb{P}\left[\bigwedge_{\mathsf{b}\in\beta}\left(\left|\widehat{x}(\mathsf{b}) - \frac{\sum\mathbf{x}}{w}\right| > \left\lceil\frac{\lg\lg w}{2}\right\rceil + 1 - \rho\right)\middle|\, \mathcal{E}\left(\mathbf{x}, \ell_\rho\right)\right] + \frac{4}{w^3} \\
<\ & w^2 \cdot \frac{4}{w^3} + \frac{4}{w^3} \\
<\ & \frac{8}{w}\,,
\end{aligned}$$

as needed. ∎

By Lemmas 3.4 and 6.5, it immediately follows:

Lemma 6.6 *Consider the group $\mathcal{L}_\rho$. Then,*

$$\begin{aligned}
& \mathbb{P}\left[\mathbf{y} \textit{ is ceiling } \left(\left\lceil\tfrac{1}{2}\lg\lg w\right\rceil + 3 - \rho\right)\textit{-concentrated }\right] \\
>\ & \mathbb{P}\left[\mathbf{x} \textit{ is ceiling } \left(\left\lceil\tfrac{1}{2}\lg\lg w\right\rceil + 3 - \rho\right)\textit{-concentrated }\right] - \frac{8}{w}\,.
\end{aligned}$$

6.4 Main Result

We show:

Theorem 6.7 *The randomized network Block_w^2 is a 17-smoothing network with probability at least* $1 - 2 \cdot \dfrac{4\lg\lg w - 39}{w}$.

Proof: We start by proving:

Lemma 6.8 *For each group $\mathcal{L}_\rho$, where $1 \leq \rho \leq \left\lceil\dfrac{\lg\lg w}{2}\right\rceil - 6$,*

$$\mathbb{P}\left[\mathbf{y}(\rho) \textit{ is ceiling } \left(\left\lceil\frac{\lg\lg w}{2}\right\rceil + 2 - \rho\right)\textit{-concentrated}\right] \ \geq\ 1 - \frac{8\rho + 1}{w}\,.$$

Proof: By induction on ρ. For the basis case where $\rho = 1$, Lemmas 6.6 and 5.2 imply that

$$\begin{aligned}
& \mathbb{P}\left[\mathbf{y}(1) \text{ is ceiling } \left(\left\lceil \frac{\lg \lg w}{2} \right\rceil + 2 - \rho\right)\text{-concentrated}\right] \\
\geq\ & \mathbb{P}\left[\mathbf{x}(1) \text{ is ceiling } \left(\left\lceil \frac{\lg \lg w}{2} \right\rceil + 2 - \rho\right)\text{-concentrated}\right] - \frac{8}{w} \\
\geq\ & \left(1 - \frac{1}{w}\right) - \frac{8}{w} \\
=\ & 1 - \frac{9}{w}\,,
\end{aligned}$$

as needed.

Assume inductively that the claim holds for $\rho - 1$, $1 < \rho \leq \left\lceil \frac{\lg \lg w}{2} \right\rceil - 6$. For the induction step, consider group $\mathcal{L}_\rho$. Then, Lemma 6.6 and the induction hypothesis imply that

$$\begin{aligned}
& \mathbb{P}\left[\mathbf{y}(\rho) \text{ is ceiling } \left(\left\lceil \frac{\lg \lg w}{2} \right\rceil + 2 - \rho\right)\text{-concentrated}\right] \\
\geq\ & \mathbb{P}\left[\mathbf{x}(\rho) \text{ is ceiling } \left(\left\lceil \frac{\lg \lg w}{2} \right\rceil + 2 - \rho\right)\text{-concentrated}\right] - \frac{1}{w} \\
\geq\ & \mathbb{P}\left[\mathbf{y}(\rho - 1) \text{ is ceiling } \left(\left\lceil \frac{\lg \lg w}{2} \right\rceil + 2 - \rho\right)\text{-concentrated}\right] - \frac{1}{w} \\
\geq\ & \left(1 - \frac{8(\rho - 1) + 1}{w}\right) - \frac{8}{w} \\
=\ & 1 - \frac{8\rho + 1}{w}\,,
\end{aligned}$$

as needed. ■

Similarly to Lemma 6.8, we prove:

Lemma 6.9 *For each group $\mathcal{L}_\rho$, where $1 \leq \rho \leq \left\lceil \frac{\lg \lg w}{2} \right\rceil - 6$,*

$$\mathbb{P}\left[\mathbf{y}(\rho) \textit{ is floor } \left(\left\lceil \frac{\lg \lg w}{2} \right\rceil + 2 - \rho\right)\textit{-concentrated}\right] \ \geq\ 1 - \frac{8\rho + 1}{w}\,.$$

Setting $\rho = \left\lceil \frac{\lg \lg w}{2} \right\rceil - 6$ in Lemma 6.8 yields that

$$\begin{aligned}
\mathbb{P}\left[\mathbf{y}\left(\left\lceil \frac{\lg \lg w}{2} \right\rceil - 6\right) \text{ is ceiling 8-concentrated}\right] \ &\geq\ 1 - \frac{8\left(\left\lceil \frac{\lg \lg w}{2} \right\rceil - 6\right) + 1}{w} \\
&\geq\ 1 - \frac{8\left(\frac{\lg \lg w}{2} + 1 - 6\right) + 1}{w} \\
&\geq\ 1 - \frac{4 \lg \lg w - 39}{w}\,.
\end{aligned}$$

Similarly, setting $\rho = \left\lceil \frac{\lg \lg w}{2} \right\rceil - 6$ in Lemma 6.9 yields that

$$\mathbb{P}\left[\mathbf{y}\left(\left\lceil\frac{\lg\lg w}{2}\right\rceil - 6\right) \text{ is floor 8-concentrated}\right] \quad \geq \quad 1 - \frac{4\lg\lg w - 39}{w}\,.$$

By Lemma 2.5, the last two inequalities imply that

$$\begin{aligned}
\mathbb{P}\left[\mathsf{CCC}_w^2(\mathbf{x}) \text{ is 17-smooth}\right] \quad &\geq \quad \mathbb{P}\left[\mathbf{y}\left(\left\lceil\frac{\lg\lg w}{2}\right\rceil - 6\right) \text{ is 17-smooth}\right] \\
&\geq \quad 1 - 2\cdot\frac{4\lg\lg w - 39}{w}\,,
\end{aligned}$$

as needed. ■

Since the cascade of two block networks is contained in the periodic network, Theorem 6.7 immediately implies:

Corollary 6.10 *The randomized periodic network is a* 17*-smoothing network with probability at least* $1 - 2\cdot\dfrac{4\lg\lg w - 39}{w}$.

7 Improbability of 1-Smoothing

We show:

Theorem 7.1 *A randomized network* B_w *is* 1*-smoothing with probability at most* $\dfrac{\mathsf{d}(B_w)}{w-1}$.

The proof will follow the *probabilistic method* [4]. Roughly speaking, we shall consider an input vector generated randomly. We shall establish that the output vector of the network B_w on the generated input vector is 1-smooth with probability at most $\dfrac{\mathsf{d}(\mathsf{B_w})}{w-1}$. This will imply the existence of an input vector such that the corresponding output vector is 1-smooth with probability at most $\dfrac{\mathsf{d}(\mathsf{B_w})}{w-1}$; from this, the claim will follow.

For any arbitrary pair of distinct integers $i, j \in [w]$, denote as $\mathbf{z}_{i,j}$ the vector

$$\mathbf{z}_{i,j} \quad = \quad \left\langle 1, \ldots, 1, \underbrace{0}_{\text{entry } i}, 1, \ldots, 1, \underbrace{2}_{\text{entry } j}, 1, \ldots, 1 \right\rangle$$

with w entries; note that the vector $\mathbf{z}_{i,j}$ is *not* 1-smooth.

For the proof, we will introduce the layer $\ell = 0$ of a balancing network as the layer preceding layer 1 with no balancers and input vector $\mathbf{x}(0) = \mathbf{x}$; so, $\mathbf{y}(0) = \mathbf{x}$ as well.

Proof: Fix a randomized 1-smoothing network B_w. Choose two distinct integers $i, j \in [w]$ uniformly at random. So, the probability that a particular pair of distinct integers $\widehat{i}$ and $\widehat{j}$ from $[w]$ are chosen is $\dfrac{1}{w(w-1)}$. Construct the input vector

$$\mathbf{x}_{i,j} \quad := \quad \mathbf{z}_{i,j};$$

clearly, $\mathbf{x}_{i,j}$ is a random variable with

$$\mathbb{P}\left[\mathbf{x}_{i,j} = \mathbf{z}_{i_0,j_0}\right] \quad = \quad \frac{1}{w(w-1)}$$

for any fixed pair of distinct wires i_0 and j_0 from $[w]$.

By the Law of Conditional Alternatives,

$$\begin{aligned}
\mathbb{P}\left[\mathsf{B}_w(\mathbf{x}_{i,j}) \text{ is 1-smooth}\right] &= \sum_{\widehat{i},\widehat{j}\in[w]\,|\,\widehat{i}\neq\widehat{j}} \mathbb{P}\left[\mathbf{x}_{i,j} = \mathbf{z}_{\widehat{i},\widehat{j}}\right] \cdot \mathbb{P}\left[\mathsf{B}_w(\mathbf{z}_{\widehat{i},\widehat{j}}) \text{ is 1-smooth}\right] \\
&= \sum_{\widehat{i},\widehat{j}\in[w]\,|\,\widehat{i}\neq\widehat{j}} \frac{1}{w(w-1)} \cdot \mathbb{P}\left[\mathsf{B}_w(\mathbf{z}_{\widehat{i},\widehat{j}}) \text{ is 1-smooth}\right] \\
&= \frac{1}{w(w-1)} \cdot \sum_{\widehat{i},\widehat{j}\in[w]\,|\,\widehat{i}\neq\widehat{j}} \mathbb{P}\left[\mathsf{B}_w(\mathbf{z}_{\widehat{i},\widehat{j}}) \text{ is 1-smooth}\right] .
\end{aligned}$$

For each layer ℓ with $1 \leq \ell \leq \mathsf{d}(\mathsf{B}_w)$, denote as $\mathcal{E}(\mathbf{x}_{i,j}, \ell)$ the event that on input vector $\mathbf{x}_{i,j}$, there is a balancer b in layer ℓ such that the two inputs to b are 0 and 2; so, by the Union Bound (Lemma 2.3),

$$\begin{aligned}
\mathbb{P}\left[\mathcal{E}(\mathbf{x}_{i,j}, \ell)\right] &= \mathbb{P}\left[\bigvee_{\mathsf{b}\in\ell} \left(\{x_1(\mathsf{b}), x_2(\mathsf{b})\} = \{0, 2\}\right)\right] \\
&\leq \sum_{\mathsf{b}\in\ell} \mathbb{P}\left[\{x_1(\mathsf{b}), x_2(\mathsf{b})\} = \{0, 2\}\right] \\
&= \sum_{\mathsf{b}\in\ell} \Big(\mathbb{P}\left[(x_1(\mathsf{b}) = 0 \wedge x_2(\mathsf{b}) = 2)\right] + \mathbb{P}\left[(x_1(\mathsf{b}) = 2 \wedge x_2(\mathsf{b}) = 0)\right]\Big) \\
&= \sum_{\mathsf{b}\in\ell} \left(\mathbb{P}\left[\mathbf{y}(\ell-1) = \mathbf{z}_{i_1(\mathsf{b}),i_2(\mathsf{b})}\right] + \mathbb{P}\left[\mathbf{y}(\ell-1) = \mathbf{z}_{i_2(\mathsf{b}),i_1(\mathsf{b})}\right]\right) .
\end{aligned}$$

Clearly, $\mathsf{B}_w(\mathbf{x}_{i,j})$ is 1-smooth if and only if there is a layer ℓ with $1 \leq \ell \leq \mathsf{d}(\mathsf{B}_w)$ such that the event $\mathcal{E}(\mathbf{x}_{i,j}, \ell)$ occurs; so, by the Union Bound (Lemma 2.3),

$$\begin{aligned}
\mathbb{P}\left[\mathsf{B}_w(\mathbf{x}_{i,j}) \text{ is 1-smooth}\right] &= \mathbb{P}\left[\bigvee_{\ell=1}^{\mathsf{d}(\mathsf{B}_w)} \mathcal{E}(\mathbf{x}_{i,j}, \ell)\right] \\
&\leq \sum_{\ell=1}^{\mathsf{d}(\mathsf{B}_w)} \mathbb{P}\left[\mathcal{E}(\mathbf{x}_{i,j}, \ell)\right] .
\end{aligned}$$

We continue with an informal outline for the rest of the proof. We shall establish that for each layer ℓ with $1 \leq \ell \leq \mathsf{d}(B_w)$, $\mathbb{P}[\mathcal{E}(\mathbf{x}_{i,j}, \ell)] \leq \frac{1}{w-1}$ (Lemma 7.3). From this fact and using the Union Bound, we will prove that $\mathbb{P}[\mathsf{B}_w(\mathbf{x}_{i,j}) \text{ is 1-smooth}] \leq \frac{\mathsf{d}(\mathsf{B}_w)}{w-1}$. Then, we will use the expression derived from the Law of Conditional Alternatives for the probability $\mathbb{P}[\mathsf{B}_w(\mathbf{x}_{i,j}) \text{ is 1-smooth}]$ to conclude the existence of a *particular* input vector $\mathbf{x}_{\widehat{i},\widehat{j}}$ such that $\mathbb{P}\left[\mathsf{B}_w(\mathbf{x}_{\widehat{i},\widehat{j}}) \text{ is 1-smooth}\right] \leq \frac{\mathsf{d}(\mathsf{B}_w)}{w-1}$; this establishes the claim. We now provide the details of the formal proof. We start with a technical claim.

Lemma 7.2 *Fix an arbitrary pair of distinct wires i_0 and j_0 from $[w]$. Then, for each layer ℓ with $0 \leq \ell \leq \mathsf{d}(\mathsf{B}_w)$,*

$$\mathbb{P}[\mathbf{y}(\ell) = \mathbf{z}_{i_0,j_0}] \quad \leq \quad \frac{1}{w(w-1)}.$$

Proof: By induction on ℓ. For the basis case, consider the layer $\ell = 0$. Clearly,

$$\begin{aligned} \mathbb{P}[\mathbf{y}(0) = \mathbf{z}_{i_0,j_0}] &= \mathbb{P}[\mathbf{x}_{i,j} = \mathbf{z}_{i_0,j_0}] \\ &= \frac{1}{w(w-1)}, \end{aligned}$$

and the claim follows.

Assume inductively that the claim holds for the layer $\ell - 1$, where $1 \leq \ell \leq \mathsf{d}(\mathsf{B}_w)$. For the induction step, consider the layer ℓ. By the Law of Conditional Alternatives and the induction hypothesis,

$$\begin{aligned} \mathbb{P}[\mathbf{y}(\ell) = \mathbf{z}_{i_0,j_0}] &= \sum_{\widehat{i},\widehat{j} \in [w] \mid \widehat{i} \neq \widehat{j}} \mathbb{P}\left[\mathbf{y}(\ell-1) = \mathbf{z}_{\widehat{i},\widehat{j}}\right] \cdot \mathbb{P}\left[\mathbf{y}(\ell) = \mathbf{z}_{i_0,j_0} \mid \mathbf{y}(\ell-1) = \mathbf{z}_{\widehat{i},\widehat{j}}\right] \\ &\leq \frac{1}{w(w-1)} \cdot \sum_{\widehat{i},\widehat{j} \in [w] \mid \widehat{i} \neq \widehat{j}} \mathbb{P}\left[\mathbf{y}(\ell) = \mathbf{z}_{i_0,j_0} \mid \mathbf{y}(\ell-1) = \mathbf{z}_{\widehat{i},\widehat{j}}\right] \\ &= \frac{1}{w(w-1)} \cdot \sum_{\widehat{i},\widehat{j} \in [w] \mid \widehat{i} \neq \widehat{j}} \mathbb{P}\left[y_{i_0}(\ell) = 0 \wedge y_{j_0}(\ell) = 2 \mid \mathbf{y}(\ell-1) = \mathbf{z}_{\widehat{i},\widehat{j}}\right]. \end{aligned}$$

We proceed by case analysis on the connections of the wires i_0 and j_0 in layer ℓ. There are five possible cases; two representative ones are illustrated in Figure 7

(C1) Assume first that i_0 is connected to some balancer $\mathsf{b} = \{i_0, i_1\}$ from layer ℓ (with $i_1 \neq i_0$), while j_0 is not connected to any balancer from layer ℓ. Clearly, for any pair of distinct wires $\widehat{i}, \widehat{j} \in [w]$,

$$\mathbb{P}\left[y_{i_0}(\ell) = 0 \wedge y_{j_0}(\ell) = 2 \mid \mathbf{y}(\ell-1) = \mathbf{z}_{\widehat{i},\widehat{j}}\right] \quad = \quad 0$$

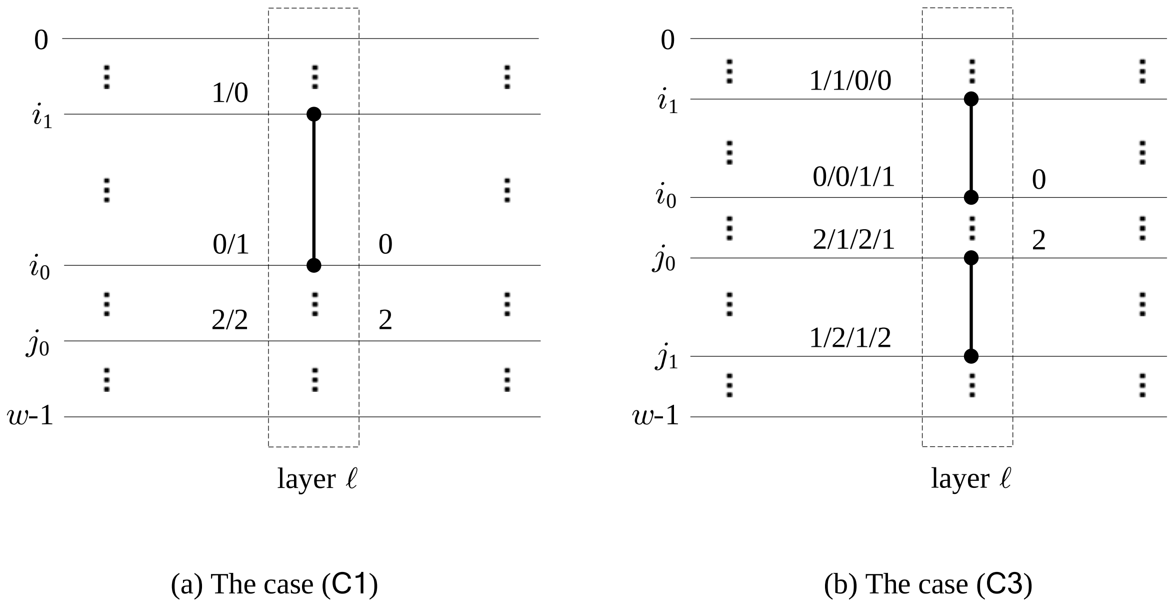

Figure 7: The cases (C1) and (C3) in the proof of Lemma 7.2. The numbers on the input and output wires of layer ℓ denote the corresponding numbers of tokens in all possible cases where $y_{i_0}(\ell) = 0$ and $y_{j_0}(\ell) = 2$.

unless $\widehat{i} \in \{i_0, i_1\}$ and $\widehat{j} = j_0$. Hence,

$$\begin{aligned}
& \sum_{\widehat{i},\widehat{j} \in [w] \,|\, \widehat{i} \neq \widehat{j}} \mathbb{P}\left[y_{i_0}(\ell) = 0 \wedge y_{j_0}(\ell) = 2 \;|\; \mathbf{y}(\ell-1) = \mathbf{z}_{\widehat{i},\widehat{j}}\right] \\
= & \sum_{\widehat{i} \in \{i_0, i_1\}, \widehat{j} = j_0} \mathbb{P}\left[y_{i_0}(\ell) = 0 \wedge y_{j_0}(\ell) = 2 \;|\; y(\ell-1) = \mathbf{z}_{\widehat{i},\widehat{j}}\right] \\
= & \sum_{\widehat{i} \in \{i_0, i_1\}, \widehat{j} = j_0} \mathbb{P}\left[y_{i_0}(\ell) = 0 \,|\; y(\ell-1) = \mathbf{z}_{\widehat{i},\widehat{j}}\right] \\
= & \;\mathbb{P}\left[\mathbf{y}_{i_0}(\ell) = 0 \;|\; \mathbf{y}(\ell-1) = \mathbf{z}_{i_0,j_0}\right] + \mathbb{P}\left[\mathbf{y}_{i_0}(\ell) = 0 \;|\; \mathbf{y}(\ell-1) = \mathbf{z}_{i_1,j_0}\right] \\
= & \;2 \cdot \mathbb{P}\left[i_0 \xrightarrow{\mathsf{b}} i_1\right] \\
= & \;2 \cdot \frac{1}{2} \\
= & \;1\,.
\end{aligned}$$

(C2) Assume now that j_0 is connected to some balancer b_j from layer ℓ, while i_0 is not connected to any balancer from layer ℓ. This case is similar to the case (C1), and we omit the analysis.

(C3) Assume now that i_0 and j_0 are connected to distinct balancers $\mathsf{b}_{i_0} = \{i_0, i_1\}$ and $\mathsf{b}_{j_0} =$

$\{j_0, j_1\}$, respectively, from layer ℓ. Clearly, for any pair of distinct wires $\widehat{\imath}, \widehat{\jmath} \in [w]$,

$$\mathbb{P}\left[y_{i_0}(\ell) = 0 \wedge y_{j_0}(\ell) = 2 \mid \mathbf{y}(\ell-1) = \mathbf{z}_{\widehat{\imath},\widehat{\jmath}}\right] \;=\; 0$$

unless $\widehat{\imath} \in \{i_0, i_1\}$ and $\widehat{\jmath} \in \{j_0, j_1\}$. Hence,

$$\begin{aligned}
& \sum_{\widehat{\imath},\widehat{\jmath} \in [w] \,|\, \widehat{\imath} \neq \widehat{\jmath}} \mathbb{P}\left[y_{i_0}(\ell) = 0 \wedge y_{j_0}(\ell) = 2 \mid \mathbf{y}(\ell-1) = \mathbf{z}_{\widehat{\imath},\widehat{\jmath}}\right] \\
= & \sum_{\widehat{\imath} \in \{i_0,i_1\}, \widehat{\jmath} \in \{j_0,j_1\}} \mathbb{P}\left[y_{i_0}(\ell) = 0 \wedge y_{j_0}(\ell) = 2 \mid \mathbf{y}(\ell-1) = \mathbf{z}_{\widehat{\imath},\widehat{\jmath}}\right] \\
= & \ \mathbb{P}\left[y_{i_0}(\ell) = 0 \wedge y_{j_0}(\ell) = 2 \mid \mathbf{y}(\ell-1) = \mathbf{z}_{i_0,j_0}\right] + \mathbb{P}\left[y_{i_0}(\ell) = 0 \wedge y_{j_0}(\ell) = 2 \mid \mathbf{y}(\ell-1) = \mathbf{z}_{i_0,j_1}\right] \\
& + \\
& \ \mathbb{P}\left[y_{i_0}(\ell) = 0 \wedge y_{j_0}(\ell) = 2 \mid \mathbf{y}(\ell-1) = \mathbf{z}_{i_1,j_0}\right] + \mathbb{P}\left[y_{i_0}(\ell) = 0 \wedge y_{j_0}(\ell) = 2 \mid \mathbf{y}(\ell-1) = \mathbf{z}_{i_1,j_1}\right] \\
= & \ \mathbb{P}\left[y_{i_0}(\ell) = 0 \mid \mathbf{y}(\ell-1) = \mathbf{z}_{i_0,j_0}\right] \cdot \mathbb{P}\left[y_{j_0}(\ell) = 2 \mid \mathbf{y}(\ell-1) = \mathbf{z}_{i_0,j_0}\right] \\
& + \mathbb{P}\left[y_{i_0}(\ell) = 0 \mid \mathbf{y}(\ell-1) = \mathbf{z}_{i_1,j_0}\right] \cdot \mathbb{P}\left[y_{j_0}(\ell) = 2 \mid \mathbf{y}(\ell-1) = \mathbf{z}_{i_1,j_0}\right] \\
& + \mathbb{P}\left[y_{i_0}(\ell) = 0 \mid \mathbf{y}(\ell-1) = \mathbf{z}_{i_0,j_1}\right] \cdot \mathbb{P}\left[y_{j_0}(\ell) = 2 \mid \mathbf{y}(\ell-1) = \mathbf{z}_{i_0,j_1}\right] \\
& + \mathbb{P}\left[y_{i_0}(\ell) = 0 \mid \mathbf{y}(\ell-1) = \mathbf{z}_{i_1,j_1}\right] \cdot \mathbb{P}\left[y_{j_0}(\ell) = 2 \mid \mathbf{y}(\ell-1) = \mathbf{z}_{i_1,j_1}\right] \\
= & \ 4 \cdot \mathbb{P}\left[i_0 \stackrel{\mathsf{b}_{i_0}}{\rightarrow} i_1\right] \cdot \mathbb{P}\left[j_0 \stackrel{\mathsf{b}_{j_0}}{\rightarrow} j_1\right] \\
= & \ 4 \cdot \frac{1}{4} \\
= & \ 1\,.
\end{aligned}$$

(C4) Assume now that i_0 and j_0 are connected to the same balancer $\mathsf{b} = \{i_0, j_0\}$ from layer ℓ. Clearly, for any pair of distinct wires $\widehat{\imath}, \widehat{\jmath} \in [w]$,

$$\mathbb{P}\left[y_{i_0}(\ell) = 0 \wedge y_{j_0}(\ell) = 2 \mid \mathbf{y}(\ell-1) = \mathbf{z}_{\widehat{\imath},\widehat{\jmath}}\right] \;=\; 0\,,$$

since the balancer property yields that $|y_{i_0}(\ell) - y_{j_0}(\ell)| \leq 1$.

(C5) Assume finally that neither i_0 nor j_0 are connected to a balancer from layer ℓ; so, $y_{i_0}(\ell) = y_{i_0}(\ell-1)$ and $y_{j_0}(\ell) = y_{j_0}(\ell-1)$. Hence,

$$\begin{aligned}
& \sum_{\widehat{\imath},\widehat{\jmath} \in [w] \,|\, \widehat{\imath} \neq \widehat{\jmath}} \mathbb{P}\left[y_{i_0}(\ell) = 0 \wedge y_{j_0}(\ell) = 2 \mid \mathbf{y}(\ell-1) = \mathbf{z}_{\widehat{\imath},\widehat{\jmath}}\right] \\
= & \sum_{\widehat{\imath},\widehat{\jmath} \in [w] \,|\, \widehat{\imath} \neq \widehat{\jmath}} \mathbb{P}\left[y_{i_0}(\ell-1) = 0 \wedge y_{j_0}(\ell-1) = 2 \mid \mathbf{y}(\ell-1) = \mathbf{z}_{\widehat{\imath},\widehat{\jmath}}\right] \\
= & \sum_{\widehat{\imath},\widehat{\jmath} \in [w] \,|\, \widehat{\imath} \neq \widehat{\jmath}} \begin{cases} 1\,, & \text{if } y_{\widehat{\imath}}(\ell-1) = 0 \text{ and } y_{\widehat{\jmath}}(\ell-1) = 2 \\ 0\,, & \text{if } y_{\widehat{\imath}}(\ell-1) \neq 0 \text{ or } y_{\widehat{\jmath}}(\ell-1) \neq 2 \end{cases} \\
\leq & \ 1\,.
\end{aligned}$$

It follows from the case analysis that

$$\mathbb{P}\left[\mathbf{y}(\ell) = \mathbf{z}_{i_0,j_0}\right] \quad \leq \quad \frac{1}{w(w-1)}\,,$$

as needed. ■

We are now ready to prove:

Lemma 7.3 *For each layer ℓ with $1 \leq \ell \leq \mathsf{d}(\mathsf{B}_w)$,*

$$\mathbb{P}\left[\mathcal{E}(\mathbf{x}_{i,j}, \ell)\right] \quad \leq \quad \frac{1}{w-1}\,.$$

Proof: By Lemma 7.2, it follows that

$$\begin{aligned}
\mathbb{P}\left[\mathcal{E}(\mathbf{x}_{i,j}, \ell)\right] \quad &\leq \quad \sum_{\mathsf{b} \in \ell} \left(\frac{1}{w(w-1)} + \frac{1}{w(w-1)} \right) \\
&\leq \quad \frac{w}{2} \cdot \frac{2}{w(w-1)} \\
&= \quad \frac{1}{w-1}\,,
\end{aligned}$$

as needed. ■

By Lemma 7.3, it follows that

$$\mathbb{P}\left[\mathsf{B}_w(\mathbf{x}_{i,j}) \text{ is 1-smooth}\right] \quad \leq \quad \frac{\mathsf{d}(\mathsf{B}_w)}{w-1}\,.$$

Hence,

$$\sum_{\widehat{i},\widehat{j} \in [w] \,|\, \widehat{i} \neq \widehat{j}} \frac{1}{w(w-1)} \cdot \mathbb{P}\left[\mathsf{B}_w(\mathbf{z}_{\widehat{i},\widehat{j}}) \text{ is 1-smooth}\right] \quad \leq \quad \frac{\mathsf{d}(\mathsf{B}_w)}{w-1}$$

or

$$\sum_{\widehat{i},\widehat{j} \in [w] \,|\, \widehat{i} \neq \widehat{j}} \mathbb{P}\left[\mathsf{B}_w(\mathbf{z}_{\widehat{i},\widehat{j}}) \text{ is 1-smooth}\right] \quad \leq \quad w \cdot \mathsf{d}(\mathsf{B}_w)\,.$$

It follows that there is an input vector $\mathbf{z}_{i_0,j_0}$ such that

$$\begin{aligned}
\mathbb{P}\left[\mathsf{B}_w(\mathbf{z}_{i_0,j_0}) \text{ is 1-smooth}\right] \quad &\leq \quad \frac{1}{w(w-1)} \cdot w \cdot \mathsf{d}(\mathsf{B}_w) \\
&= \quad \frac{\mathsf{d}(\mathsf{B}_w)}{w-1}\,.
\end{aligned}$$

Hence, B_w is a 1-smoothing network with probability at most $\frac{\mathsf{d}(\mathsf{B}_w)}{w-1}$, as needed. ■

Theorem 7.1 immediately implies:

Corollary 7.4 *There is no small-depth, randomized* 1*-smoothing network with constant probability.*

8 Conclusion

We presented a thorough study of the impact of randomization in smoothing networks. We showed a *tight* (up to a small additive constant) bound of $\lg \lg w + \Theta(1)$ on the smoothness of the randomized block network. As our main result, we established an upper bound of 17 on the smoothness of the cascade of two randomized block networks. Finally, we established that it is impossible to obtain a randomized 1-smoothing network of small depth and sufficiently high probability.

These results delimit the power of randomization in smoothing networks: it can be employed in a simple, small-depth network to yield a constant upper bound on smoothness; however, it *cannot* be employed to yield an upper bound of 1 on smoothness unless the network has linear depth.

Our work leaves open numerous interesting problems. The most interesting extension of our research is to establish the conjecture that the cascade of a (small) constant number of randomized blocks is a 2-smoothing network (with high probability). (By Theorem 7.1, this is the best trade-off one could hope for.) More generally, it would be very interesting to establish trade-offs between the number of cascaded (randomized) block networks and the achievable smoothness.

On a more concrete level, it would be interesting to close the small gap between the constants (-2 and 3) from the lower and upper bounds on the smoothness of the (randomized) block. Also, how tight is the upper bound of 17 on the smoothness of the cascade of two randomized blocks? Finally, how tight is the improbability result from Theorem 7.1? Is there a *specific* network B_w which, on some particular input vector, achieves smoothness of 1 with probability *exactly* $\frac{\mathsf{d}(\mathsf{B}_w)}{w-1}$?

Acknowledgments.

We thank Maurice Herlihy and Srikanta Tirthapura whose work [13] has inspired ours. We also thank Costas Busch and the anonymous *PODC 2008* reviewers for their helpful comments.

References

[1] E. Aharonson and H. Attiya, "Counting Networks with Arbitrary Fan-Out," *Distributed Computing,* Vol. 8, No. 4, pp. 163–169, 1995.

[2] W. Aiello, R. Venkatesan and M. Yung, "Coins, Weights and Contention in Balancing Networks," *Proceedings of the 13th Annual ACM Symposium on Principles of Distributed Computing,* pp. 193–205, August 1994.

[3] M. Ajtai, J. Komlós and E. Szemerédi, "Sorting in $c \lg n$ Parallel Steps," *Combinatorica,* Vol. 3, No. 1, pp. 1–19, 1983.

[4] N. Alon and J. Spencer, *The Probabilistic Method,* Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, Second Edition, 2000.

[5] J. Aspnes, M. Herlihy and N. Shavit, "Counting Networks," *Journal of the ACM,* Vol. 41, No. 5, pp. 1020–1048, September 1994.

[6] K. E. Batcher, "Sorting Networks and Their Applications," *Proceedings of the AFIPS Joint Computer Conference,* pp. 334–338, 1968.

[7] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods,* Athena Scientific, 1997.

[8] C. Busch and M. Mavronicolas, "An Efficient Counting Network," *Proceedings of the 1st Merged International Parallel Processing Symposium and IEEE Symposium on Parallel and Distributed Processing,* pp. 380–384, May 1998.

[9] M. Dowd, Y. Perl, L. Rudoplh and M. Saks, "The Periodic Balanced Sorting Network," *Journal of the ACM,* Vol. 36, No. 4, pp. 738–757, October 1989.

[10] C. Dwork, M. Herlihy and O. Waarts, "Contention in Shared Memory Algorithms," *Journal of the ACM,* Vol. 44, No. 6, pp. 779–805, November 1997.

[11] B. Fristedt and L. Gray, *A Modern Approach to Probability Theory,* Probability and Its Applications, Birkhäuser, 1997.

[12] M. Herlihy and N. Shavit, *The Art of Multiprocessor Programming,* Morgan Kaufmann/Elsevier, 2008.

[13] M. Herlihy and S. Tirthapura, "Randomized Smoothing Networks," *Journal of Parallel and Distributed Computing,* Vol. 66, No. 5, pp. 626–632, May 2006.

[14] M. Herlihy and S. Tirthapura, "Self-Stabilizing Smoothing and Counting Networks," *Distributed Computing,* Vol. 18, No. 5, pp. 345–357, April 2006.

[15] W. Hoeffding. "Probability Inequalities for Sums of Bounded Random Variables," *Journal of the American Statistical Association,* Vol. 53, pp. 13–30, 1963.

[16] S. Kapidakis and M. Mavronicolas, "Distributed, Low Contention Task Allocation," *Proceedings of the 8th IEEE Symposium on Parallel and Distributed Processing,* pp. 358–365, October 1996.

[17] M. Klugerman, *Small-Depth Counting Networks and Related Topics,* Ph.D. Thesis, Department of Mathematics, Massachusetts Institute of Technology, September 1994.

[18] M. Klugerman and C. G. Plaxton, "Small-Depth Counting Networks," *Proceedings of the 24th Annual ACM Symposium on Theory of Computing,* pp. 417–428, May 1992.

[19] D. E. Knuth, *The Art of Computer Programming, Vol. 3/Sorting and Searching*, Second Edition, Addison-Wesley, 1998.

[20] C. P. Kruskal and M. Snir, "A Unified Theory of Interconnection Network Structure," *Theoretical Computer Science,* Vol. 48, No. 3, pp. 75–94, 1986.

[21] F. Preparata and J. Vuillemin, "The Cube-Connected-Cycles: A Versatile Network for Parallel Computation," *Communications of the ACM.* Vol. 24, No. 5, pp. 300–309, May 1981.

[22] M. D. Riedel and J. Bruck, "Tolerating Faults in Counting Networks," Technical Report ETR-22, Paradise, California Institute of Technology, April 1999.