The Impact of Randomization in Smoothing Networks*

Marios Mavronicolas
Department of Computer Science
University of Cyprus
CY-1678 Nicosia, Cyprus
mavronic@cs.ucy.ac.cy

Thomas Sauerwald
Department of Computer Science
University of Paderborn
D-33102 Paderborn, Germany
sauerwal@upb.de

ABSTRACT

We revisit *smoothing networks* [3], which are made up of *balancers* and *wires*. *Tokens* arrive arbitrarily on w *input wires* and propagate asynchronously through the network; each token gets service 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 assume that balancers are oriented independently and uniformly at random. We present a collection of lower and upper bounds on smoothness, which are to some extent surprising:

- The smoothness of a single *block network* [7] is $\lg \lg w + \Theta(1)$ (with high probability), where the additive constant is between -2 and 4. This *tight* bound improves vastly over the upper bound of $\mathcal{O}(\sqrt{\lg w})$ from [9], and it significantly improves 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 16 (with high probability); this is the *first* known randomized network with so small depth ($2 \lg w$) and so good smoothness. The proof introduces some novel combinatorial and probabilistic structures and techniques which may be further applicable. This 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* 1-smoothing network in [14], this result implies the *first* separation between deterministic and randomized smoothing networks, which demonstrates an unexpected limitation of randomization: it can get to constant smoothness very easily, but after that, the progress to 1-smoothing is very limited.

Category and Subject Descriptors: G.3 [Probability and Statistics]: Stochastic Processes - *Probabilistic algorithms*; C.2.4. [Distributed Systems]: Distributed Applications - *Load Balancing*
General Terms: Algorithms, Theory
Keywords: Smoothing networks, Randomization

1. INTRODUCTION

A *smoothing network* [3] is a distributed data structure that receives *tokens* issued arbitrarily by multiple concurrent processes at *input wires* and routes them asynchronously through a network to *output wires*. The network consists of switches (called *balancers*), and *wires*. A balancer is oriented either top or bottom; the first token through the balancer will be forwarded to its (local) *top* or *bottom* output wire, respectively, and subsequent tokens will alternate. Each token represents a request by a *client* for a service to unit work; the service is provided by the *server* residing on the output wire the token arrives at. Tokens are dispersed through the network, thereby reducing *contention* (cf. [3]).

The routing of tokens through the network must ensure that all servers receive approximately the same number of tokens, no matter how unbalanced arrival of tokens on input wires is. The *smoothness* of a smoothing network is the maximum discrepancy among the numbers of tokens arriving at different output wires; a γ-*smoothing network* has smoothness γ. Smoothing networks with low smoothness are attractive for multiprocessor coordination and load balancing applications where low-contention is a requirement; these include *producers-consumers* [8] and distributed numerical computations [6]. Together with *counting networks*, smoothing networks have been studied quite extensively since introduced in the seminal paper of Aspnes *et al.* [3]. It has been a major open problem to construct efficient and small-depth (counting and) smoothing networks (cf. [5, 14]).

We only require *local initialization* for the balancers of a smoothing network (cf. [1, 9, 10]). Specifically, we assume that each balancer is oriented either top or bottom uniformly at random during some initialization phase. Although local initialization withdraws the advantage of global consistency offered by *global initialization*, it is still attractive since it offers fault-tolerance against crashes, resets or replacements of balancers. Such smoothing networks were called *randomized smoothing networks* [1, 9].

Herlihy and Tirthapura [9] studied the smoothing prop-

*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)

PODC'08, August 18–21, 2008, Toronto, Ontario, Canada.

erties of the *block network* Block_w [3, 7] when all balancers are initialized uniformly at random; the block network is a very simple network of depth $\lg w$ that has been used in advanced constructions such as the *periodic (counting) network* [3, 7]. An upper bound of $2.36\sqrt{\lg w}$ (with high probability) was shown in [9]; this bound is trivially inherited to the *bitonic network* [3, 4] and the periodic network [3, 7] since they both contain the block. The upper bound from [9] improved vastly over the smoothness of $\lg w$ known before for simple constructions (such as the *bitonic merger* [12] and the *butterfly* [13, 14]) with global initialization, and for the block network itself with local (*arbitrary* and not randomized) initialization [10]. Klugerman [13] and Klugerman and Plaxton [14] had earlier presented an elaborate construction of a network with smoothness 1; however, their network is impractical since it contains the AKS network [2] inheriting huge constants.

Herlihy and Tirthapura formulated three interesting *Open Problems* about randomized smoothing networks in [9]:

1. Our bounds for the smoothness of the block network does 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?

In this work, we provide answers to all these problems. We first prove that Block_w is $(\lg\lg w + 4)$-smoothing with probability at least $1 - 4w^{-3}$ (Theorem 5.5). Our proof drastically improves the elementary techniques developed by Herlihy and Tirthapura [9] for their corresponding proof of the $\mathcal{O}(\sqrt{\lg w})$ upper bound. The exponential improvement is achieved through a tighter analysis of the same random variables. In more detail, we provide a judicious partition of the block network into two groups of layers, and we analyze separately the influence of each group on smoothness; the first group consists of $\lg w - \lg\lg w$ layers. Hence, a certain sum of independent random variables that is bounded using a Chernoff bound in [9] is now split into two sums; one of them is again bounded with a Chernoff bound, while the other is bounded deterministically by simply summing the maximum possible absolute values of the random variable. This result provides a partial answer to *Open Problem 3* of Herlihy and Tirthapura [9].

We continue to establish a matching lower bound (up to a small additive constant) on the smoothness of the block network. More precisely, we prove that Block_w is only a $(\lg\lg w - 2)$-smoothing network with probability at most $2\exp(-\frac{4\sqrt{w}}{\lg w})$ (Theorem 5.6); thus, the analysis is essentially tight. The proof again partitions the network into two groups of layers. We determine a *fixed point* input for the first group; we then prove that (with high probability) this input is not smoothed better than $\lg\lg w - 1$ when traversing the second group. This result completes the answer to *Open Problem 3* of Herlihy and Tirthapura [9].

As our main result we show that there is a very simple and shallow, randomized network that is $\mathcal{O}(1)$-smoothing. Specifically, consider the cascade of two block networks; we prove that the cascade is 16-smoothing with probability at least $1 - \frac{4\lg\lg w - 47}{w}$ (Theorem 6.1). The proof uses a judicious partition of the second block network into no more than $\frac{1}{2}\lg\lg w - 6$ groups of layers; the number of layers per group increases as we proceed. We show that each group drops the smoothness by 1. Hence, at the end, the application of Theorem 5.5 to the first cascaded block network implies a constant smoothness. To establish each drop of smoothness by 1, we employ some very delicate probabilistic arguments; we believe that these will be useful elsewhere — for example, in showing that a small number (greater than 2) of cascaded block networks is 2-smoothing (with high probability), and we conjecture this to be the case.

We remark that our result on the smoothness of the cascade of two block networks provides an answer to *Open Problem 1* of Herlihy and Tirthapura [9]: When the input to a block network has the properties of the output of a block (for example, it is $(\lg\lg w + 4)$-smooth), then its output is 16-smooth (with high probability). Since the cascade of two block networks is contained in the periodic network [3, 7] (which consists of $\lg w$ such blocks), the latter is also 16-smoothing. This settles *Open Problem 2* of Herlihy and Tirthapura [9]. Finally, we note that this result identifies the *first* $\Theta(\lg w)$-depth (randomized) smoothing network that simultaneously *(i)* achieves constant smoothness, *(ii)* does not use the AKS network [2] (and, hence, it need not be inpractical) and *(iii)* does not require global initialization.

We conclude with an improbability result: There is no randomized network of width w and depth d that achieves 1-smoothing with probability greater than $\frac{\mathsf{d}}{w-1}$ (Theorem 7.1). This is bad news: It implies that the output of any of the common (randomized) networks of depth $\mathcal{O}(\lg^2 w)$ (such as the periodic network [7]) is 1-smooth with an extremely small probability. Furthermore, only randomized smoothing networks of depth *linear* in w may guarantee 1-smoothness with high probability; so, it is impossible to obtain a shallow 1-smoothing network through randomization. Since there is a deterministic 1-smoothing network (relying, however, on the AKS network to achieve depth $\Theta(\lg w)$) [13, 14], this result provides the *first* separation between deterministic and randomized (1-)smoothing networks, and demonstrates a somehow unexpected limit on the impact of randomization in smoothing networks: there is some c between 1 and 15 such that there are shallow randomized $(c+1)$-smoothing networks with high probability, but no shallow randomized c-smoothing networks. The proof establishes that on a certain random input, the output is 1-smooth with probability at most $\frac{\mathsf{d}}{w-1}$. This implies the existence of a *deterministic* input with this property, which implies the claim.

2. PRELIMINARIES AND NOTATION

All logarithms are to the base 2. For an integer x, the binary representation of x is a binary word $x_1x_2\ldots x_k$ with $k \geq \lg x$ such that $\sum_{i=1}^{k} 2^{k-i}x_i = x$. For an integer $x \geq 1$, denote $[x] = \{0, \ldots, x-1\}$. For a number $x \in \mathbb{R}$, denote $\exp(x) = e^x$. For an integer $x \geq 0$, the *odd-characteristic* function of x, denoted as $\mathsf{Odd}(x)$, is given by $\mathsf{Odd}(x) = 1$ if x is odd, and 0 otherwise.

We denote by $\mathbf{x}$ a vector $\langle x_0, \ldots, x_{w-1}\rangle$ of w integers. For a vector $\mathbf{x}$, denote $\sum\mathbf{x} = \sum_{i\in[w]} x_i$. Say that $\mathbf{x}$ is ***γ-smooth*** if for every pair x_i, x_j, $|x_i - x_j| \leq \gamma$. We shall use the ***Hoeffding Bound*** [11]:

LEMMA 2.1 (HOEFFDING BOUND). *Let $v_i \in [a_i, b_i]$, $i \in$*

Network	Depth	type	GI?	Smoothness	Probability	Reference
KP network	$\Theta(\lg w)$	D	✓	1	1	[13, 14]
r-butterfly	$\lg w + o(\lg w)$	D/R	✓	2	$1 - \frac{1}{\text{superpoly}(w)}$	[1]
Bitonic merger	$\lg w$	D	✓	$\lg w$	1	[12]
Butterfly	$\lg w$	D	✓	$\lg w$	1	[14]
Block	$\lg w$	D	X	$\lg w$	1	[10, Theorems 3 & 4]
Block	$\lg w$	R	X	$2.36\sqrt{\lg w}$	$1 - 4w^{-1}$	[9, Theorem 10]
Block	$\lg w$	R	X	$\lg\lg w + 4$	$1 - 4w^{-3}$	Theorem 5.5
Block	$\lg w$	R	X	$\lg\lg w - 2$	$\leq 2\exp(-\frac{4\sqrt{w}}{\lg w})$	Theorem 5.6
Two Blocks	$2\lg w$	R	X	16	$1 - \frac{4\lg\lg w - 47}{w}$	Theorem 6.1
Any network	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; D/R stands for a combination of deterministic and randomized balancers. GI stands for global initialization; that column indicates whether GI is required or not. KP stands for Klugerman and Plaxton [13, 14]. By way of comparison, the KP network uses huge constant factors in its depth due to its reliance on the AKS network, while there are no hidden constants in our 16-smoothing network of two blocks. Futhermore, some balancers in the r-butterfly network [1] require deterministic initialization in order to achieve 2-smoothness.

$[n]$, *be independent random variables. Then, for* $\delta \geq 0$, $\mathbb{P}\left[\left|\sum_{i=1}^{n} v_i - \mathbf{E}\left[\sum_{i=1}^{n} v_i\right]\right| \geq \delta\right] \leq 2\cdot\exp\left(-\frac{2\delta^2}{\sum_{i=1}^{n}(b_i-a_i)^2}\right)$.

Our coming proofs will not consider the possibility that certain expressions (e.g., $\lg\lg w$ and $\frac{\sum \mathbf{x}}{w}$) may not be integer. Adding floors and ceilings in the analysis will suffice to address the general case. For the seek of notational simplicity, we have opted to present the simpler analysis with no floors and ceilings.

3. SMOOTHING NETWORKS

A ***smoothing network*** [3] is a special case of a ***balancing network*** [3], which is a collection of interconnected ***balancers***. A balancer is an asynchronous switch with two ***input wires*** and two ***output wires***, called ***top*** and ***bottom***. An ***initialization*** takes places in some preprocessing phase; the initialization simply chooses an ***orientation***: one of the two output wires, either the top or the bottom. The balancer is oriented top (resp. bottom) if the initialization chooses its top (resp. bottom) output wire. A stream of tokens enters a balancer via its two input wires; each time a new token arrives on an input wire, it is directed to the output wire currently labeled top; at the same time, the orientation of the balancer changes. This ensures that the total number of tokens is (almost) evenly divided among the two output wires.

A ***balancing network*** 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 connected to any input wires. When the numbers of input and output wires of the network are the same, this number w is called the ***width*** of the network and the network is denoted by B_w. The acyclicity ensures that each balancer can be assigned a unique ***layer***: the length of the longest path from an input wire to that layer; the ***depth*** $\mathsf{d}(\mathsf{B}_w)$ is the maximum layer.

The network $\mathsf{Prefix}_\ell(\mathsf{B}_w)$ consists of the layers $1, \ldots, \ell$ of B_w; the network $\mathsf{Suffix}_\ell(\mathsf{B}_w)$ consists of the layers $\mathsf{d}(\mathsf{B}_w) - \ell + 1, \ldots, \mathsf{d}(\mathsf{B}_w)$. Finally, for an integer $k \geq 1$, B_w^κ denotes the sequential cascade of κ copies of B_w. Say that a balancer b in layer ℓ of a balancing network B ***depends*** on balancer b' in layer $\ell' < \ell$ if there is a path from b' to b in B. Then, each output wire of balancer b depends on balancer b' as well (and also trivially 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'. Consider two output wires j_1 and j_2 of layer ℓ in a balancing network B. 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'.

We make a distinction according to the way balancers are initialized. A ***deterministic balancer*** is one that is initialized in some deterministic way. A ***deterministic balancing network*** consists of deterministic balancers. A pair of a deterministic balancing network B_w and a (fixed) orientation for each of its balancers induces a set of (asynchronous) ***executions*** in the natural way (cf. [3, Section 2]). Consider an ***input vector*** $\mathbf{x} = \langle x_0, x_1, \ldots, x_{w-1}\rangle$, where x_i is the number of tokens fed into input wire i of the network B_w. A ***quiescent state*** of the network B_w on the input vector $\mathbf{x}$ is reached in some execution when all $\sum \mathbf{x}$ input tokens have exited. It is simple to observe that *all* executions of network B_w (on the input vector $\mathbf{x}$) reach a quiescent state with a common output vector $\mathbf{y} = \langle y_0, y_1, \ldots, y_{w-1}\rangle$. So, identify each quiescent state with the 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. [10]).

Say that B_w is a γ-***smoothing network*** for some integer $\gamma \geq 1$ (possibly dependent on w) if for each input vector $\mathbf{x}$, $\mathsf{B}_w(\mathbf{x})$ is γ-smooth.

A ***randomized balancer*** [1, 9] is initialized to each of top and bottom with probability $\frac{1}{2}$ and independently of all other balancers; so, it is initialized uniformly at random. A ***randomized balancing network*** consists of randomized balancers. So, a randomized balancing network is a pair of a balancing network B_w and a random orientation of it. Given an input vector $\mathbf{x}$ to a randomized balancing network, induced in the natural way is a probability measure $\mathbb{P}$ on associated events. 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 $\mathbf{x}$, $\mathbb{P}\left[\mathsf{B}_w(\mathbf{x}) \text{ is } \gamma\text{-smooth}\right] \geq \delta$; that is, the probability that for all output wires j and k, $1 \leq j, k \leq w$, it holds that $|y_j - y_k| \leq \gamma$ is at least δ.

For a balancer b denote as x_1 and x_2 the number of tokens arriving on the top and bottom input wires of b, respectively. Denote as y_1 and y_2 the number of tokens leaving through the top and bottom output wires of b, respectively. (We shall sometimes use $x_1(\mathsf{b}), x_2(\mathsf{b}), y_1(\mathsf{b}), y_2(\mathsf{b})$ for x_1, x_2, y_1 and y_2, respectively, when reference to b is necessary.) If b is oriented top (resp., bottom), $y_1 = \left\lceil \frac{x_1+x_2}{2} \right\rceil$ 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 $y_2 = \left\lceil \frac{x_1+x_2}{2} \right\rceil$). Assume now that b is oriented uniformly at random. Define a random variable r_{b} taking values $\frac{1}{2}$ and $-\frac{1}{2}$ with equal probability (cf. [9]). (Clearly, $\mathbb{E}[r_{\mathsf{b}}] = 0$.) Define $x_{\mathsf{b}} = \mathsf{Odd}(x_1 + x_2) \cdot r_{\mathsf{b}}$ (cf. [9]). Then, $y_1 = \frac{x_1+x_2}{2} + x_{\mathsf{b}} = \frac{x_1+x_2}{2} + \mathsf{Odd}(x_1+x_2) \cdot r_{\mathsf{b}}$ and $y_1 = \frac{x_1+x_2}{2} - x_{\mathsf{b}} = \frac{x_1+x_2}{2} - \mathsf{Odd}(x_1+x_2) \cdot r_{\mathsf{b}}$.

4. BLOCK NETWORK (AND RELATIVES)

The Block_2 network is a single balancer. The Block_{2w} is constructed from two Block_w networks as follows. Given an input vector $\mathbf{x}^{(2w)}$, represent each subscript as a binary string. The ***A-cochain*** of $\mathbf{x}^{(2w)}$, denoted as $\mathbf{x}_A$, is the subvector whose indices have low-order bits 00 or 11; the ***B-cochain*** of $\mathbf{x}^{(2w)}$ of $\mathbf{x}^{(2w)}$, denoted as $\mathbf{x}_B$, is the subvector whose indices have low-order bits 01 or 10. The input vector $\mathbf{x}^{(2w)}$ is fed into two parallel Block_w networks, so that $\mathbf{x}_A$ and $\mathbf{x}_B$ go to each of the Block_w networks. Denote as $\mathbf{y}_A$ and $\mathbf{y}_B$ the corresponding outputs of the two Block_w networks. In a final layer, each pair of corresponding entries of $\mathbf{y}_A$ and $\mathbf{y}_B$ are matched through a balancer. So, Block_w has $\lg w$ layers $1, \ldots, \lg w$, each with $\frac{w}{2}$ balancers (cf. Figure 1).

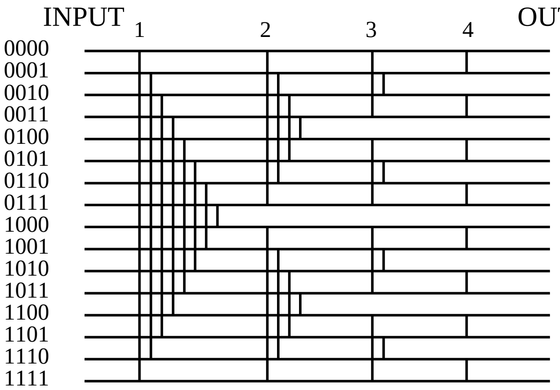

Figure 1: The Block_{16} network.

We will use the tree structure from [9, Section 2] for Block_w:

- The *root* is the set of all $\frac{w}{2}$ balancers at layer 1 of Block_w; label this node $v_{1,1}$. The *leaves* of the tree are the balancers in layer $\lg w$.
- For each ℓ, $2 \leq \ell \leq \lg w$, layer ℓ is decomposed into $2^{\ell-1}$ nodes, denoted as $v_{\ell,1}, \ldots, v_{\ell,2^{\ell-1}}$, each consisting of $\frac{w}{2^\ell}$ balancers. These nodes are defined inductively (given the nodes for layer $\ell - 1$): For each integer k, where $1 \leq k \leq 2^{\ell-2}$, the node $v_{\ell,2k-1}$ consists of all balancers (in layer ℓ) that the top output wires of balancers in node $v_{\ell-1,k}$ point to. Similarly, the node $v_{\ell,2k}$ consists of all balancers (in layer ℓ) which the bottom output wires of balancer in node $v_{\ell-1,k}$ point to.

The tokens that exit from output wire y_1 must follow the path $v_{1,1}, v_{2,1}, \ldots, v_{\lg w,1}$ and exit on the top output wire of balancer $v_{\lg w,1}$. See Figure 2 for an illustration of the tree structure. We observe a preliminary property of Block_w, easily shown by induction, which will be used later.

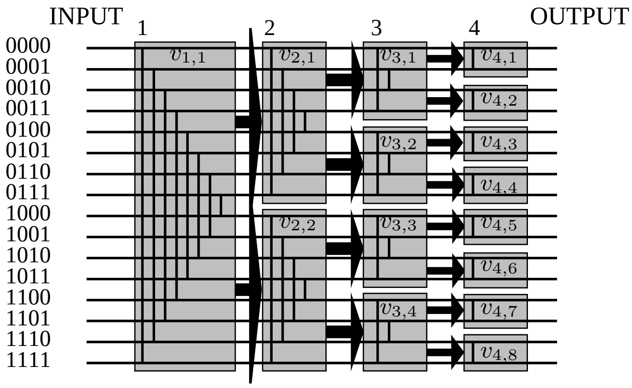

Figure 2: The tree structure of [9] within a Block_{16}.

LEMMA 4.1. *In Block_w, there is at most one path from a balancer* b *in layer ℓ to a balancer* b$'$ *in layer $\ell' > \ell$.*

The block network is very similar to (but different than) the well-known ***merger*** network of Batcher [4]. In more detail, under the *standard* orientation (cf. [9]), there is no permutation between the input wires of the two networks that yields one from the other while respecting the orientation of each balancer. However, if the balancers' orientations are ignored, such permutations exist and the networks are called ***isomorphic*** (cf. [7, Section 2]). The $\mathsf{Periodic}_w$ network is the cascade of $\lg w$ Block_w networks.

Cube-Connected-Cycles: For w a power of 2, the network CCC_w has $\lg w$ layers. In layer ℓ, $1 \leq \ell \leq \lg w$, for each wire $u \in \{0,1\}^{\lg w}$, there is a balancer b between wire u and wire $u(\ell)$, where $u(\ell) = u_1 \ldots u_{\ell-1}\overline{u_\ell}u_{\ell+1} \ldots u_{\lg w}$; the top output wire of b is the one among u and $u(\ell)$ such that $u_\ell = 0$. See Figure 3 for an illustration. We observe a simple structural property of CCC_w.

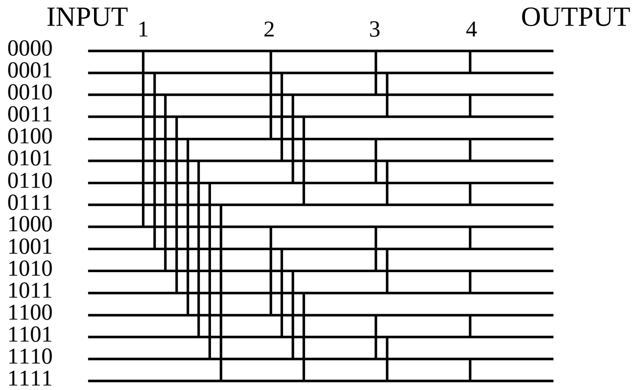

Figure 3: The CCC_{16} network.

LEMMA 4.2. *Consider two integers l_1 and l_2 such that $l_1 + l_2 < \lg w$. Let $\ell_1 \in \{0,1\}^{l_1}$ and $\ell_2 \in \{0,1\}^{l_2}$ be arbitrary but fixed. Then, the restriction of CCC_w to the layers $l_1 + 1, \ldots, \lg w - l_2$ and wires $\ell_1 u \ell_2$, where $u \in \{0,1\}^{\lg w - l_1 - l_2}$ is arbitrary, is a network $\mathsf{CCC}_{2^{\lg w - l_1 - l_2}}$.*

It is simple to see that the block network is a ***bidelta*** network [15] (A bidelta network is a one that is ***delta*** network in both directions (from left to right and vice versa); roughly speaking, a delta network is one in which there is a unique path from each input wire to every output wire, and the path descriptors associated with paths leading to the same output wire are identical.) The cube-connected-cycles is another example of a bidelta network. It is known that any two bidelta networks of the same width (and degree 2, say) are isomorphic [15]. Hence, the block network is isomorphic to the cube-connected-cycles network (ignoring the balancer orientations). This allows to treat the two networks interchangeably when considering random orientations.

The numbers $y_0, y_1, \ldots, y_{w-1}$ of tokens on output wires $0, 1, \ldots, w-1$ are random variables of the block network with random orientation. The symmetry of the block network implies that all variables $y_j, 0 \leq j \leq w-1$, follow the same distribution (cf. [9]).

For a layer ℓ and an integer j, where $1 \leq j \leq 2^{\ell-1}$, denote as $x_{\ell,j}$ and $y_{\ell,j}$ the total number of tokens entering and exiting node $v_{\ell,j}$, respectively. Since the tokens on the top (resp. bottom) output wires of the balancers in $v_{\ell,j}$ enter the node $v_{\ell+1,2j-1}$ (resp. $v_{\ell+1,2j}$), the numbers of tokens entering nodes $v_{\ell+1,2j-1}$ and $v_{\ell+1,2j}$ are $x_{\ell+1,2j-1} = \frac{x_{\ell,j}}{2} + \sum_{\mathsf{b}\in v_{\ell,j}} x_{\mathsf{b}}$ and $x_{\ell+1,2j} = \frac{x_{\ell,j}}{2} - \sum_{\mathsf{b}\in v_{\ell,j}} x_{\mathsf{b}}$, respectively. Since all random variables $y_j, 0 \leq j \leq w-1$, follow the same distribution, we focus on the number of tokens y_0 exiting on the top output wire of node $v_{\lg w,1}$. To calculate y_0, we need to count the number of tokens following the path $v_{1,1}, v_{2,1}, \ldots, v_{\lg w,1}$ and exiting on the top output wire of $v_{\lg w,1}$. We recall:

LEMMA 4.3 (HERLIHY AND TIRTHAPURA [9]). *For the network* Block_w, $y_0 = \frac{\sum \mathbf{x}}{w} + \sum_{\ell=1}^{\lg w} \frac{1}{2^{\lg w - \ell}} \sum_{\mathsf{b}\in v_{\ell,1}} x_{\mathsf{b}}$.

LEMMA 4.4 (HERLIHY AND TIRTHAPURA [9]). *Let* $\mathcal{B}$ *be a set of balancers in* Block_w *and* c_{b} *be a constant for each balancer* $\mathsf{b} \in \mathcal{B}$. *Then, for any* $\delta > 0$, $\mathbb{P}\left[\sum_{\mathsf{b}\in\mathcal{B}} c_{\mathsf{b}} x_{\mathsf{b}} > \delta\right] \leq 2 \cdot \mathbb{P}\left[\sum_{\mathsf{b}\in\mathcal{B}} c_{\mathsf{b}} r_{\mathsf{b}} > \delta\right]$.

Note that $\sum_{\mathsf{b}\in\mathcal{B}} c_{\mathsf{b}} x_{\mathsf{b}}$ is a sum of *dependent* random variables, while $\sum_{\mathsf{b}\in\mathcal{B}} c_{\mathsf{b}} r_{\mathsf{b}}$ is a sum of *independent* random variables. Furthermore, note that for each set of balancers $\mathcal{B}$, linearity of expectations implies $\mathbb{E}\left[\sum_{\mathsf{b}\in\mathcal{B}} c_{\mathsf{b}} r_{b}\right] = 0$. The following claim can be derived easily using techniques from [9].

LEMMA 4.5. *For* Block_w, $\mathbb{P}\left[\left|y_0 - \frac{\sum \mathbf{x}}{w}\right| \geq \delta\right] \leq 4\exp(-\delta^2)$.

PROOF. First, we observe the following intermediate technical equality, which was also observed in [9, page 5]:

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

By Lemma 4.4,

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

For each pair of layer $\ell, 1 \leq \ell \leq \lg w$ and a balancer $\mathsf{b} \in v_{\ell,1}$, the variable r_{b} has range $\{-2^{-\lg w + \ell - 1}, 2^{-\lg w + \ell - 1}\}$. By Hoeffding Bound (Lemma 2.1) and the observed inequality, we get $\mathbb{P}\left[\left|\sum_{\ell=1}^{\lg w} \frac{1}{2^{\lg w - \ell}} \sum_{\mathsf{b}\in v_{\ell,1}} x_{\mathsf{b}}\right| \geq \delta\right] \leq 4\exp(-\delta^2)$. □

5. ONE BLOCK

We present both upper and lower bounds on the smoothness of Block_w. We start with some observations. By Lemma 4.3,

$$y_0 = \frac{\sum \mathbf{x}}{w} + \sum_{\ell=1}^{\lg w} \frac{1}{2^{\lg w - \ell}} \sum_{\mathsf{b}\in v_{\ell,1}} x_{\mathsf{b}} = \frac{\sum \mathbf{x}}{w} + \underbrace{\sum_{\ell=1}^{\lg w - \lg\lg w} \frac{1}{2^{\lg w - \ell}} \sum_{\mathsf{b}\in v_{\ell,1}} x_{\mathsf{b}}}_{X_1} + \underbrace{\sum_{\ell=\lg w - \lg\lg w + 1}^{\lg w} \frac{1}{2^{\lg w - \ell}} \sum_{\mathsf{b}\in v_{\ell,1}} x_{\mathsf{b}}}_{X_2},$$

where for each layer ℓ, $|v_{\ell,1}| = 2^{\lg w - \ell}$. We first prove:

LEMMA 5.1. *For the network* Block_w, $\sum_{\ell=1}^{\lg w - \lg\lg w} \sum_{\mathsf{b}\in v_{\ell,1}} \left(\frac{1}{2^{\lg w - \ell + 1}} - \left(-\frac{1}{2^{\lg w - \ell + 1}}\right)\right)^2 = \frac{2}{\lg w} - \frac{2}{w}$.

Now we prove two preliminary properties of X_1 and X_2.

LEMMA 5.2. $\mathbb{P}\left[|X_1| \geq 2\right] \leq 4w^{-4}$

PROOF. Define $R_1 = \sum_{\ell=1}^{\lg w - \lg\lg w} \frac{1}{2^{\lg w - \ell}} \sum_{\mathsf{b}\in v_{\ell,1}} r_{\mathsf{b}}$. By Lemma 4.4, $\mathbb{P}\left[|X_1| \geq 2\right] \leq 2 \cdot \mathbb{P}\left[|R_1| \geq 2\right]$. For each pair of a layer ℓ, $1 \leq \ell \leq \lg w - \lg\lg w$, and a balancer $\mathsf{b} \in v_{\ell,1}$, the variable $\frac{1}{2^{\lg w - \ell}} \cdot r_{\mathsf{b}}$ has range $\{-2^{-\lg w + \ell - 1}, +2^{-\lg w + \ell - 1}\}$. By Hoeffding Bound and Lemma 5.1, $\mathbb{P}\left[|X_1| \geq 2\right] \leq 2 \cdot 2 \cdot \exp\left(-\frac{2 \cdot 2^2}{\frac{2}{\lg w}}\right) \leq 4w^{-4}$. □

We now use the triangle inequality to prove:

LEMMA 5.3. $|X_2| \leq \frac{1}{2}\lg\lg w$.

Lemmas 5.2 and 5.3, and the Union Bound imply:

LEMMA 5.4. $\mathbb{P}\left[\bigvee_{k\in[w]} \left(\left|y_k - \frac{\sum \mathbf{x}}{w}\right| \geq \frac{1}{2}\lg\lg w + 2\right)\right] \leq 4w^{-3}$.

We start with the upper bound:

THEOREM 5.5. Block_w *is a* $(\lg\lg w + 4)$*-smoothing network with probability at least* $1 - 4w^{-3}$.

PROOF. The event $\bigwedge_{k\in[w]} \left(\left|y_k - \frac{\sum \mathbf{x}}{w}\right| < \frac{1}{2}\lg\lg w + 2\right)$ implies that for each pair of indices $k, l \in [w]$, $|y_k - y_l| \leq \left|y_k - \frac{\sum \mathbf{x}}{w}\right| + \left|-y_l + \frac{\sum \mathbf{x}}{w}\right| \leq \frac{1}{2}\lg\lg w + 2 + \frac{1}{2}\lg\lg w + 2 = \lg\lg w + 4$. By Lemma 5.4,

$$\begin{aligned} & \mathbb{P}\left[\mathsf{B}_w(\mathbf{x}) \text{ is } (\lg\lg w + 4)\text{-smooth}\right] \\ =\ & \mathbb{P}\left[\bigwedge_{k,l\in[w]} \left(|y_k - y_l| \leq \lg\lg w + 4\right)\right] \\ \geq\ & \mathbb{P}\left[\bigwedge_{k\in[w]} \left(\left|y_k - \frac{\sum \mathbf{x}}{w}\right| \leq \frac{1}{2}\lg\lg w + 2\right)\right] \\ \geq\ & 1 - \mathbb{P}\left[\bigvee_{k\in[w]} \left(\left|y_k - \frac{\sum \mathbf{x}}{w}\right| \geq \frac{1}{2}\lg\lg w + 2\right)\right] \\ \geq\ & 1 - 4w^{-3}. \end{aligned}$$

□

We continue with the lower bound:

THEOREM 5.6. Block_w *is a* $(\lg\lg w - 2)$*-smoothing network with probability at most* $2\exp(-\frac{4\sqrt{w}}{\lg w})$.

Since the networks Block_w and CCC_w are isomorphic, we shall deal in the proof with the second.

PROOF. We construct an input vector $\mathbf{x}$ such that the probability that $\mathbf{y} = \mathsf{CCC}_w(\mathbf{x})$ is $(\lg\lg w - 2)$-smooth is at most $2\exp(-\frac{4\sqrt{w}}{\lg w})$. Construct $\mathbf{x}$ as follows. For each input wire $i = i_1 i_2 \ldots i_{\lg w}$, set $x_i := \sum_{k=\lg w - \lg\lg w + 2}^{\lg w} i_k$; so, x_i is the number of 1's in the $\lg\lg w - 1$ least significant bits of $i_1 i_2 \ldots i_{\lg w}$ (An illustration is given in Figure 4). We prove:

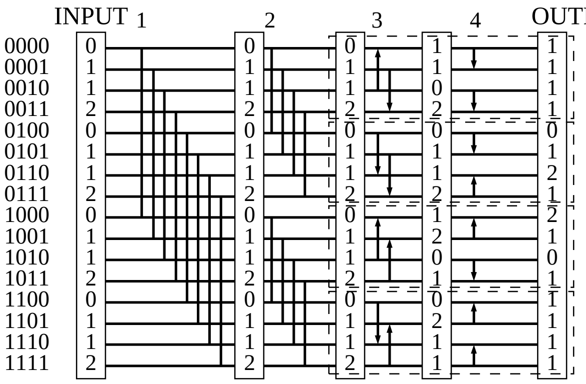

Figure 4: The input vector for the lower bound together with a sample initialization of the CCC_{16}.

LEMMA 5.7. $\mathbf{x}$ *is fixed point of* $\mathsf{Prefix}_{\lg w - \lg\lg w + 1}(\mathsf{CCC}_w)$.

PROOF. We prove that for each layer $\ell \leq \lg w - \lg\lg w + 1$, $\mathbf{x}$ is a fixed point of $\mathsf{Prefix}_\ell(\mathsf{CCC}_w)$. The proof is by induction on ℓ. For the basis case, where $\ell = 1$, consider a balancer b in layer ℓ connecting wires u and $u(1)$, where $u \in \{0,1\}^{\lg w}$. By construction of $\mathbf{x}$, the inputs to balancer b are

$$x_1(\mathsf{b}) = \sum_{k=\lg w - \lg\lg w + 2}^{\lg w} u_k, \;\; x_2(\mathsf{b}) = \sum_{k=\lg w - \lg\lg w + 2}^{\lg w} u(1)_k.$$

By construction of the CCC_w, $u_k = u(1)_k$ for all $k \geq \lg w - \lg\lg w + 2$ (since u and $u(1)$ differ only in bit 1.) Hence, $x_1(\mathsf{b}) = x_2(\mathsf{b})$. By definition of balancers, $y_1(\mathsf{b}) = y_2(\mathsf{b})$. Hence, $\mathbf{x}$ is a fixed point of $\mathsf{Prefix}_\ell(\mathsf{CCC}_w)$.

Assume inductively that the claim holds for layer $\ell - 1$, where $1 < \ell < \lg w - \lg\lg w + 2$. Consider a balancer b in layer ℓ connecting wires u and $u(\ell)$, where $u \in \{0,1\}^{\lg w}$. By induction hypothesis, $x_1(\mathsf{b}) = \sum_{k=\lg w - \lg\lg w + 2}^{\lg w} u_k$ and $x_2(\mathsf{b}) = \sum_{k=\lg w - \lg\lg w + 2}^{\lg w} u(\ell)_k$. By construction of the network, $u_k = u(\ell)_k$ for all $k \geq \lg w - \lg\lg w + 2$ (since u and $u(\ell)$ differ only in bit ℓ.) Hence, $x_1(\mathsf{b}) = x_2(\mathsf{b})$. By definition of balancers, it follows that $y_1(\mathsf{b}) = y_2(\mathsf{b})$. Hence, $\mathbf{x}$ is a fixed point of $\mathsf{Prefix}_\ell(\mathsf{CCC}_w)$, and the claim follows. □

We now focus on $\mathsf{Suffix}_{\lg\lg w - 1}(\mathsf{CCC}_w)$ which is, by construction, the parallel cascade of $\frac{2w}{\lg w}$ $\mathsf{CCC}_{\frac{\lg w}{2}}$ networks. Take any such network $\mathsf{CCC}_{\frac{\lg w}{2}}$. The input wires of such a CCC_w are $u0^{\lg\lg w - 1}, \ldots, u1^{\lg\lg w - 1}$, where $u \in \{0,1\}^{\lg w - \lg\lg w + 1}$; by Lemma 5.7, the input to $i = uv$, where $v \in \{0,1\}^{\lg\lg w - 1}$, is $\sum_{k=\lg w - \lg\lg w + 2}^{\lg w} i_k$. We prove:

LEMMA 5.8. $\mathbb{P}\left[y_{u1^{\lg\lg w - 1}} = 0\right] \geq 2^{-(\frac{\lg w}{2} - 1)}$ *and* $\mathbb{P}\left[y_{u1^{\lg\lg w - 1}} = \lg\lg w - 1\right] \geq 2^{-(\frac{\lg w}{2} - 1)}$.

PROOF. Note that the output wire $u1^{\lg\lg w - 1}$ depends on $1 + \sum_{k=1}^{\lg\lg w - 2} 2^k = \frac{\lg w}{2} - 1$ balancers in layers $\lg w - \lg\lg w + 1, \ldots, \lg w$. Notice also that there are $2^{\frac{\lg w}{2} - 1}$ orientations for these balancers, each occuring with the same probability. Hence, it suffices to prove that each of 0 and $\lg\lg w - 1$ is a possible output for the output wire $u1^{\lg\lg w - 1}$.

For simplicity, set $w' = \lg w$. The proof is by induction on w'. For the basis case, where $w' = 4$, the claim is verified directly (see also Figure 3). Assume inductively that the output wire $u1^{\lg w'}$ in the network $\mathsf{CCC}_{w'}$ can take the values 0 and $\lg w'$. For the induction step, consider the network $\mathsf{CCC}_{2w'}$. Consider the output wire $0u1^{\lg w'}$. By construction of the Cube-Connected-Cycles network, $\mathsf{CCC}_{2w'}$ consists of a ladder network followed by two parallel $\mathsf{CCC}_{w'}$ networks. Consider the top of these $\mathsf{CCC}_{w'}$ networks.

- Assume that all balancers in layer 1 of the $\mathsf{CCC}_{2w'}$ are initialized bottom. Then, the input to each of the input wires of $\mathsf{CCC}_{w'}$ equals the number of 1's in the corresponding input wire $0i'$, where $i' \in \{0,1\}^{\lg(2w') - 1} = \{0,1\}^{\lg w'}$. Clearly, this number equals the number of 1's in the string i'. Induction hypothesis implies that the output wire $u1^{\lg w'}$ can have value 0.
- Assume now that all balancers in layer 1 of the $\mathsf{CCC}_{2w'}$ are initialized top. Then, the input to each of the input wires of CCC'_w equals the number of 1's in the corresponding input wire $i = 1i'$ where $i' \in \{0,1\}^{\lg w'}$. Clearly, this number equals 1 plus the number of 1's in the string i'. Induction hypothesis implies that the output wire $u1^{\lg w'}$ can have output $1 + \lg w' = \lg 2w'$.

The proof is now complete. □

Consider two different subnetworks $\mathsf{CCC}_{\frac{\lg w}{2}}$ with input wires $u0^{\lg\lg w - 1}, \ldots u1^{\lg\lg w - 1}$ and $u'0^{\lg\lg w - 1}, \ldots, u'1^{\lg\lg w - 1}$, respectively. Consider output wires $u1^{\lg\lg w - 1}$ and $u'1^{\lg\lg w - 1}$, respectively. We now prove, using the structure of the CCC_w:

LEMMA 5.9. *The set* $\{y_{u1^{\lg\lg w - 1}} \mid u \in \{0,1\}^{\lg w - \lg\lg w + 1}\}$ *is a set of independent random variables.*

Now, by Lemmas 5.8 and 5.9, we get that

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

Similarly, we obtain that

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

By the Union Bound, it follows that

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

6. TWO BLOCKS

We consider the cascade of two networks Block_w. The analysis of the smoothness properties of the second cascaded block will require some smoothness properties of the first one which go beyond the one stated in Theorem 5.5. We show:

THEOREM 6.1. *For w sufficiently large, Block_w^2 is a 16-smoothing network with probability at least $1 - \frac{4\lg\lg w - 47}{w}$.*

Here is an informal outline of the proof. Recall that by Theorem 5.5, the first Block_w is $(\lg\lg w + 4)$-smoothing with high probability. The proof will use no more than $\frac{\lg\lg w}{2} - 6$ phases; in phase ρ, $1 \leq \rho \leq \frac{\lg\lg w}{2} - 6$, we consider a distinct group of $\frac{4\lg w}{(\lg\lg w - 1 - \rho)^2} + \lg\lg w$ consecutive layers in the second Block_w. For each phase $\rho > 1$, we prove that the smoothness of the cascade of the first Block_w and the layers considered in the phase ρ drops by one over the corresponding smoothness for phase $\rho - 1$. At the end, this will establish that the smoothness will become constant. A diagram for the proof appears in Figure 5. The formal proof follows.

PROOF. Since Block_w and CCC_w are isomorphic, we will deal with the cascade of two networks CCC_w. Consider the second cascaded CCC_w with layers $1, 2, \ldots, \lg w$. We prove:

LEMMA 6.2. *Fix a layer ℓ, with $\lg w + 1 \leq \ell \leq 2\lg w$, in the network CCC_w^2. Consider the input vector $\mathbf{x}$ to layer ℓ. Then, with probability $1 - 4w^{-3}$, the following condition holds: For every ζ, $0 \leq \zeta \leq \lg w - \lg\lg w$, and for all pairs $\ell_1 \in \{0,1\}^{\ell-1-\lg w}$ and $\ell_2 \in \{0,1\}^{2\lg w - \ell + 1 - \lg\lg w - \zeta}$,*

$$\left| \frac{\sum_{u\in\{0,1\}^{\lg\lg w+\zeta}} x_{\ell_1 u \ell_2}}{\lg w \cdot 2^{\zeta}} - \frac{\sum \mathbf{x}}{w} \right| \leq 2.$$

PROOF. We first observe:

LEMMA 6.3.

$$\sum_{u\in\{0,1\}^{\lg\lg w+\zeta}} \; \sum_{k=\ell-1-\lg w+\lg\lg w+\zeta}^{\ell-1} \; \sum_{\substack{\mathsf{b}_k\in k \;\mid\; \ell_1 u \ell_2 \\ \text{depends on } \mathsf{b}_k}} \left(\tfrac{1}{\lg w}2^{k-\ell-\zeta} - (-\tfrac{1}{\lg w}2^{k-\ell-\zeta})\right)^2 \leq \tfrac{2}{\lg w}.$$

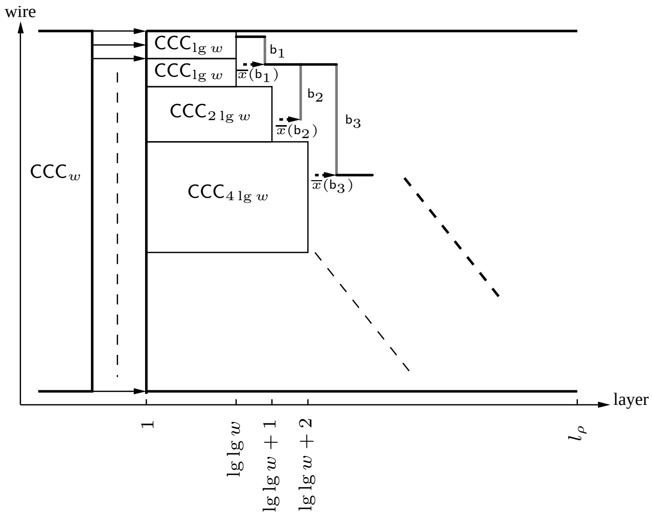

Figure 5: A diagram for the proof for the cascade of two CCC_w in group $\mathcal{L}_1$. The inputs $\overline{x}(\mathsf{b}_1), \overline{x}(\mathsf{b}_2), \overline{x}(\mathsf{b}_3), \ldots$ on the path π from the top input wire in layer $\lg\lg w$ to some output wire in layer l_ρ are all independent (for fixed input at layer 1) and are output wires of a $\mathsf{CCC}_{\lg w}, \mathsf{CCC}_{2\lg w}, \mathsf{CCC}_{4\lg w}, \ldots$.

For some $i \in \{0,1\}^{\lg w}$, let $x_i(\ell)$ be the input to wire i in layer $1 \leq \ell \leq \lg w$ in CCC_w^2. We establish that

$$\sum_{u\in\{0,1\}^{\lg\lg w+\zeta}} x_{\ell_1 u \ell_2} = \frac{1}{2^{\lg w - \lg\lg w - \zeta}} \cdot \sum \mathbf{x} + \sum_{u\in\{0,1\}^{\lg\lg w+\zeta}} \; \sum_{\substack{k=\ell-1-\lg w+ \\ \lg\lg w+\zeta+1}}^{\ell-1} 2^{k-(\ell-1)} \sum_{\substack{\mathsf{b}_k\in k:\; \ell_1 u \ell_2 \\ \text{dep. on } \mathsf{b}_k}} x_{\mathsf{b}_k}.$$

Hence,

$$\frac{\sum_{u\in\{0,1\}^{\lg\lg w+\zeta}} x_{\ell_1 u \ell_2}}{\lg w \cdot 2^{\zeta}} = \frac{\sum \mathbf{x}}{w} + \frac{1}{\lg w \cdot 2^{\zeta}} \cdot \sum_{u\in\{0,1\}^{\lg\lg w+\zeta}} \; \sum_{\substack{k=\ell-1-\lg w+ \\ \lg\lg w+\zeta+1}}^{\ell-1} 2^{k-(\ell-1)} \sum_{\substack{\mathsf{b}_k\in k:\; \ell_1 u \ell_2 \\ \text{dep. on } \mathsf{b}_k}} x_{\mathsf{b}_k}$$

Define

$$X = \frac{1}{\lg w \cdot 2^{\zeta}} \sum_{u\in\{0,1\}^{\lg\lg w+\zeta}} \; \sum_{\substack{k=\ell-1-\lg w+ \\ \lg\lg w+\zeta+1}}^{\ell-1} 2^{k-(\ell-1)} \sum_{\substack{\mathsf{b}_k\in k \;\mid\; \ell_1 u \ell_2 \\ \text{dep. on } \mathsf{b}_k}} x_{\mathsf{b}_k},$$

$$R = \frac{1}{\lg w \cdot 2^{\zeta}} \sum_{u\in\{0,1\}^{\lg\lg w+\zeta}} \; \sum_{\substack{k=\ell-1-\lg w+ \\ \lg\lg w+\zeta+1}}^{\ell-1} 2^{k-(\ell-1)} \sum_{\substack{\mathsf{b}_k\in k \;\mid\; \ell_1 u \ell_2 \\ \text{dep. on } \mathsf{b}_k}} r_{\mathsf{b}_k}.$$

By Lemma 4.4, $\mathbb{P}[|X| \geq 2] \leq 2\cdot\mathbb{P}[|R| \geq 2]$. For each pair of a $u \in \{0,1\}^{\lg\lg w+\zeta}$ and a layer k, where $\ell - 1 - \lg w + \lg\lg w + \zeta + 1 \leq k \leq \ell - 1$, such that $\ell_1 u \ell_2$ depends on some balancer $\mathsf{b}_k \in k$, the variable $\frac{1}{\lg w\cdot 2^{\zeta}} \cdot 2^{k-(\ell-1)} \cdot r_{\mathsf{b}_k}$ has range $\{-\frac{1}{\lg w}2^{k-(\ell-1)-\zeta}\cdot\frac{1}{2}, +\frac{1}{\lg w}2^{k-(\ell-1)-\zeta}\cdot\frac{1}{2}\}$

$= \{-\frac{1}{\lg w}2^{k-\ell-\zeta}, +\frac{1}{\lg w}2^{k-\ell-\zeta}\}$. By Hoeffding Bound and Lemma 6.3, $\mathbb{P}[|X| \geq 2] \leq 2\cdot 2\cdot \exp\left(-\frac{2\cdot 2^2}{\frac{2}{\lg w}}\right) \leq 4w^{-4}$. Define $\mathcal{C} := \{0,1\}^{\lg w - \lg\lg w - \zeta}$. By the Union Bound,

$$\mathbb{P}\left[\bigvee_{\zeta=0}^{\lg w - \lg\lg w}\bigvee_{\ell_1\ell_2\in\mathcal{C}}\left|\frac{\sum_{u\in\{0,1\}^{\lg\lg w+\zeta}} x_{\ell_1 u \ell_2}}{\lg w \cdot 2^{\zeta}} - \frac{\sum \mathbf{x}}{w}\right| \leq 2\right]$$
$$\leq \sum_{\zeta=0}^{\lg w - \lg\lg w}\sum_{\ell_1,\ell_2\in\mathcal{C}}\mathbb{P}\left[\left|\frac{\sum_{u\in\{0,1\}^{\lg\lg w+\zeta}} x_{\ell_1 u \ell_2}}{\lg w \cdot 2^{\zeta}} - \frac{\sum \mathbf{x}}{w}\right| \leq 2\right]$$
$$\leq \lg w \cdot 2^{\ell-1-\lg w}\cdot 2^{2\lg w-\ell+1-\lg\lg w-\zeta}\cdot\frac{4}{w^4} \leq \frac{4}{w^3}. \qquad \square$$

Consider again the second cascaded CCC_w with layers $1, 2, \ldots, \lg w$. Consider groups of layers $\mathcal{L}_1, \mathcal{L}_2, \ldots, \mathcal{L}_{\frac{\lg\lg w}{2}-6}$ in this network, defined inductively as follows (see Figure 5):

- Basis case: $\mathcal{L}_1$ has layers $1, 2, \ldots, \frac{4\lg w}{(\frac{1}{2}\lg\lg w-2)^2}+\lg\lg w$.
- Assume inductively that we have defined $\mathcal{L}_{\rho-1}$, $\rho > 2$.
- Induction step: $\mathcal{L}_\rho$ consists of the $\frac{4\lg w}{(\frac{1}{2}\lg\lg w-1-\rho)^2}+\lg\lg w$ layers which immediately follow group $\mathcal{L}_{\rho-1}$.

Denote as ℓ_ρ the first layer in group $\mathcal{L}_\rho$. By simple calculations, we obtain:

LEMMA 6.4. *For sufficiently large w, the total number of layers of groups $\mathcal{L}_1, \mathcal{L}_2, \ldots, \mathcal{L}_{\frac{\lg\lg w}{2}-6}$ is at most $\lg w$.*

Consider a path $\pi = \mathsf{b}_1, \mathsf{b}_2, \ldots, \mathsf{b}_k$ (of balancers). For each balancer b_r, $1 < r \leq k$, $x(\mathsf{b}_r)$ is the input to balancer b_r from balancer b_{r-1} and $\overline{x}_{\mathsf{b}_r}$ is the other input to balancer b_r. ($x(\mathsf{b}_1)$ is arbitrarily chosen among $x_1(\mathsf{b}_1)$ and $x_2(\mathsf{b}_1)$.) We now prove a key claim:

LEMMA 6.5. *Consider a path $\pi = \mathsf{b}_1, \mathsf{b}_2, \ldots, \mathsf{b}_{\mathsf{d}(\mathcal{L}_\rho)-\lg\lg w}$ from an input wire in layer $\lg\lg w$ of group $\mathcal{L}_\rho$ to an output wire of $\mathcal{L}_\rho$. Then,*

$$\mathbb{P}\left[\bigvee_{\mathsf{b}\in\pi}\left|\overline{x}(\mathsf{b}_r) - \frac{\sum\mathbf{x}}{w}\right| \leq \frac{1}{2}\lg\lg w + 1 - \rho\right] \leq 1 - 8\cdot w^{-3}.$$

PROOF. We first prove a technical claim we need:

CLAIM 6.6. *Consider an input vector $\mathbf{x}$ to $\mathcal{L}_\rho$. Then, the random variables in the set $\{\overline{x}(\mathsf{b}_1), \ldots, \overline{x}(\mathsf{b}_{\mathsf{d}(\mathcal{L}_\rho)-\lg\lg w})\}$ are independent.*

PROOF. Each variable $\overline{x}(\mathsf{b}_r)$, where $1 \leq r \leq \mathsf{d}(\mathcal{L}_\rho) - \lg\lg w$ is determined by *(i)* the inputs to the balancers in layer 1 of $\mathcal{L}_\rho$ on which b_r depends, and *(ii)* the (randomly) chosen orientation of the balancers of $\mathcal{L}_\rho$ on which b_r depends. By Lemma 4.1, the dependency sets of $\mathsf{b}_r, 1 \leq r \leq d(\mathcal{L}_\rho) - \lg\lg w$, are disjoint and the claim follows. $\square$

We continue with a second technical claim:

CLAIM 6.7. *Fix $\ell_2 \in \{0,1\}^{\lg w - l_\rho - \lg\lg w - \zeta + 1}$ with $\zeta \geq 0$ and fix $\ell_1 \in \{0,1\}^{\lg w - 1}$. Consider an input vector $\boldsymbol{x}$ to $\mathcal{L}_\rho$ such that $\left|\frac{\sum_{u\in\{0,1\}^{\lg\lg w+\zeta}} x_{\ell_1 u\ell_2}}{2^\zeta\cdot\lg w} - \frac{\sum\mathbf{x}}{w}\right| \leq 2$. Then, for each balancer b_r, $1 \leq r \leq \mathsf{d}(\mathcal{L}_\rho) - \lg\lg w$,*

$$\mathbb{P}\left[\left|\overline{x}(\mathsf{b}_r) - \frac{\sum\mathbf{x}}{w}\right| \geq \frac{1}{2}\lg\lg w + 1 - \rho\right] \leq 4\cdot\exp\left(-\left(\frac{1}{2}\lg\lg w - 1 - \rho\right)^2\right).$$

PROOF. Fix a balancer b_r, where $1 \leq r \leq \mathsf{d}(\mathcal{L}_\rho) - \lg\lg w$ in layer $1 \leq \ell(r) \leq \lg w$. Let $i = \overline{\ell_1(r)u(r)\ell_2(r)}$ and $i(\ell)$ be the input wires of b_r, where $\overline{\ell_1(r)} \in \{0,1\}^{\ell_\rho-1}$, $\overline{u(r)} \in \{0,1\}^{\lg\lg w+r}$ and $\overline{\ell_2(r)} \in \{0,1\}^{\lg w-\ell_\rho-\lg\lg w-r+1}$. Consider the restriction of group $\mathcal{L}_\rho$ to layers $\ell_\rho, \ell_\rho+1, \ldots, \ell_\rho+\lg\lg w+r-1$ and wires $\overline{\ell_1(r)}u\overline{\ell_2(r)}$, where $u \in \{0,1\}^{\lg\lg w+r}$. Lemma 4.2 implies that this restriction is a $\mathsf{CCC}_{2^r\lg w}$. Hence, $\overline{x}(\mathsf{b}_r)$ is some output of the network $\mathsf{CCC}_{2^r\lg w}$; notice that the input vector to this network comes from the (arbitrary but fixed) input vector $\mathbf{x}$ to $\mathcal{L}_\rho$. We finally use the triangle inequality, the assumption and Lemma 4.5 to prove that

$$\mathbb{P}\left[\left|\overline{x}(\mathsf{b}_r) - \frac{\sum\mathbf{x}}{w}\right| \geq \frac{1}{2}\lg\lg w + 1 - \rho\right]$$
$$\leq \mathbb{P}\left[\left|\overline{x}(\mathsf{b}_r) - \frac{\sum_{u\in\{0,1\}^{\lg\lg w+r}} x_{\overline{\ell_1}u\overline{\ell_2}}}{2^r\lg w}\right| + \left|\frac{\sum_{u\in\{0,1\}^{\lg\lg w+r}} x_{\overline{\ell_1}u\overline{\ell_2}}}{2^r\lg w} - \frac{\sum\mathbf{x}}{w}\right| \geq \frac{1}{2}\lg\lg w + 1 - \rho\right]$$
$$\leq \mathbb{P}\left[\left|\overline{x}(\mathsf{b}_r) - \frac{\sum_{u\in\{0,1\}^{\lg\lg w+r}} x_{\overline{\ell_1}u\overline{\ell_2}}}{2^r\lg w}\right| + 2 \geq \frac{1}{2}\lg\lg w + 1 - \rho\right]$$
$$= \mathbb{P}\left[\left|\overline{x}(\mathsf{b}_r) - \frac{\sum_{u\in\{0,1\}^{\lg\lg w+r}} x_{\overline{\ell_1}u\overline{\ell_2}}}{2^r\lg w}\right| \geq \frac{1}{2}\lg\lg w - 1 - \rho\right]$$
$$\leq 4\cdot\exp\left(-\left(\frac{1}{2}\lg\lg w - 1 - \rho\right)^2\right). \qquad \square$$

We continue with the proof of the claim. Denote by $\mathcal{E}$ the event that $\forall\ \zeta \geq 0, \ell_1 \in \{0,1\}^{\ell_\rho}$, $\ell_2 \in \{0,1\}^{\lg w - \ell_\rho - \lg\lg w - \zeta}$, $\left|\frac{\sum_{u\in\{0,1\}^{\lg\lg w+\zeta}} x_{\ell_1 u\ell_2}}{2^\zeta\cdot\lg w} - \frac{\sum\mathbf{x}}{w}\right| \leq 2$. Define $\alpha := \frac{1}{2}\lg\lg w + 1 - \rho$. By Claims 6.6 and 6.7 and Lemma 6.2,

$$\mathbb{P}\left[\bigwedge_{b\in\pi}\left|\overline{x}(\mathsf{b}_r) - \frac{\sum\mathbf{x}}{w}\right| \geq \alpha\right]$$
$$= \sum_{\substack{\mathbf{x}(\ell_\rho):\\ \text{ind. by }\mathbf{x}}}\mathbb{P}\left[\bigwedge_{b\in\pi}\left|\overline{x}(\mathsf{b}_r) - \frac{\sum\mathbf{x}}{w}\right| \geq \alpha \mid \mathbf{x}(\ell_\rho)\right] \cdot \mathbb{P}\left[\mathbf{x}(\ell_\rho)\right]$$

$$\leq \sum_{\substack{\mathbf{x}(\ell_\rho):\\ \text{ind. by } \mathbf{x} \text{ and } \mathcal{E}}} \prod_{r=1}^{\substack{\mathsf{d}(\mathcal{L}_\rho)-\\ \lg\lg w}} \mathbb{P}\left[\left|\overline{x}(\mathsf{b}_r) - \frac{\sum \mathbf{x}}{w}\right| \geq \alpha \;\middle|\; x(\ell_\rho)\right] \cdot \mathbb{P}\left[\mathbf{x}(\ell_\rho)\right] + \sum_{\substack{\mathbf{x}(\ell_\rho):\\ \text{ind. by } \mathbf{x} \text{ and } \neg\mathcal{E}}} 1 \cdot \mathbb{P}\left[\mathbf{x}(\ell_\rho)\right]$$

$$\leq \sum_{\substack{\mathbf{x}(\ell_\rho):\\ \text{ind. by } \mathbf{x} \text{ and } \mathcal{E}}} \prod_{r=1}^{\substack{\mathsf{d}(\mathcal{L}_\rho)-\\ \lg\lg w}} 4\exp\left(-(\alpha-2)^2\right) \cdot \mathbb{P}\left[\mathbf{x}(\ell_\rho)\right] + \frac{4}{w^3}$$

$$\leq 4 \cdot \exp\left(-(\mathsf{d}(\mathcal{L}_\rho) - \lg\lg w) \cdot (\alpha-2)^2\right) \cdot 1 + \frac{4}{w^3}$$

$$= 4 \cdot \exp\left(-\frac{4\lg w}{(\alpha-2)^2} \cdot (\alpha-2)^2\right) + \frac{4}{w^3} \leq \frac{8}{w^3}. \qquad \square$$

We continue to prove:

Lemma 6.8. *Consider the cascade of all layers from ℓ to ℓ', $\ell' > \ell$, with input vector $\mathbf{x}$ and output vector $\mathbf{y}$, respectively. Assume that (i) for every $i \in [w]$, $|x_i - \frac{\sum \mathbf{x}}{w}| \leq \gamma$ and (ii) for every path $\pi = \mathsf{b}_\ell, \mathsf{b}_{\ell+1}, \ldots, \mathsf{b}_{\ell'}$ from layer ℓ to layer ℓ', there is at least one layer r, with $\ell \leq r \leq \ell'$, such that $|\overline{x}(\mathsf{b}_r) - \frac{\sum \mathbf{x}}{w}| \leq \gamma - 2$. Then, $|y_i - \frac{\sum \mathbf{x}}{w}| \leq \gamma - 1$ for every $i \in [w]$.*

Proof. By contradiction. Assume that there is an $i \in [w]$ such that $|y_i - \frac{\sum \mathbf{x}}{w}| = \gamma$. Without loss of generality, assume that $y_i \geq \frac{\sum \mathbf{x}}{w} + \gamma$. Let $\mathsf{b}_{\ell'}$ be the balancer in layer ℓ' with output wire y_i. By definition of the balancer and assumption on the input vector x in layer $\ell \leq \ell'$, the two inputs must satisfy $x_1(\mathsf{b}_r) \geq \gamma + \frac{\sum \mathbf{x}}{w} - 1$ and $x_2(\mathsf{b}_r) \geq \frac{\sum \mathbf{x}}{w} + \gamma$ for an arbitrary ordering of the two input wires of b_r. Hence, there must be a path $\pi = \mathsf{b}_\ell, \mathsf{b}_{\ell+1}, \ldots, \mathsf{b}_{\ell'}$ such that for all $\ell \leq r \leq \ell'$, $|\overline{x}(\mathsf{b}_r) - \frac{\sum \mathbf{x}}{w}| \geq \gamma - 1$. A contradiction. $\square$

We are now ready to prove:

Lemma 6.9. *For an integer ρ, where $1 \leq \rho \leq \frac{\lg\lg w}{2} - 6$, consider the input and output vectors $\mathbf{x}(\rho)$ and $\mathbf{y}(\rho)$ respectively, to group $\mathcal{L}_\rho$. Then,*

$$\mathbb{P}\left[\bigwedge_{k\in[w]}\left(\left|y_k(\rho) - \frac{\sum \mathbf{x}}{w}\right| \leq \frac{1}{2}\lg\lg w + 2 - \rho\right)\right] \geq \mathbb{P}\left[\bigwedge_{k\in[w]}\left(\left|x_k(\rho) - \frac{\sum \mathbf{x}}{w}\right| \leq \frac{1}{2}\lg\lg w + 3 - \rho\right)\right] - \frac{8}{w}.$$

Proof. Let $\mathcal{P}$ be the set of all paths from the first layer of $\mathcal{L}_\rho$ to the last layer of $\mathcal{L}_\rho$. Clearly, $|\mathcal{P}| \leq w \cdot 2^{\mathsf{d}(\mathcal{L}_\rho)} \leq w \cdot 2^{\lg w} \leq w^2$. Hence, by the Union Bound and Lemma 6.5,

$$\mathbb{P}\left[\bigwedge_{\pi\in\mathcal{P}}\left(\bigvee_{\mathsf{b}\in\pi}\left|\overline{x}(\mathsf{b}_r) - \frac{\sum \mathbf{x}}{w}\right| \leq \frac{1}{2}\lg\lg w + 1 - \rho\right)\right]$$
$$\geq 1 - \sum_{\pi\in\mathcal{P}} \mathbb{P}\left[\left(\bigwedge_{\mathsf{b}\in\pi}\left|\overline{x}(\mathsf{b}_r) - \frac{\sum \mathbf{x}}{w}\right| > \frac{1}{2}\lg\lg w + 1 - \rho\right)\right]$$
$$= 1 - w^2 \cdot 8w^{-3} = 1 - 8w^{-1}.$$

By Lemma 6.8,

$$\left(\bigwedge_{k\in[w]}\left(\left|x_k - \frac{\sum \mathbf{x}}{w}\right| \leq \frac{1}{2}\lg\lg w + 3 - \rho\right)\right) \bigwedge \left(\bigwedge_{\pi\in\mathcal{P}}\left(\bigvee_{\mathsf{b}\in\pi}\left|\overline{x}(\mathsf{b}_r) - \frac{\sum \mathbf{x}}{w}\right| \leq \frac{1}{2}\lg\lg w + 1 - \rho\right)\right)$$
$$\Rightarrow \bigwedge_{k\in[w]}\left(\left|y_k - \frac{\sum \mathbf{x}}{w}\right| \leq \frac{1}{2}\lg\lg w + 2 - \rho\right).$$

So, by the Union Bound,

$$\mathbb{P}\left[\bigwedge_{k\in[w]}\left(\left|y_k - \frac{\sum \mathbf{x}}{w}\right| \leq \frac{1}{2}\lg\lg w + 2 - \rho\right)\right]$$
$$\geq \mathbb{P}\left[\left(\bigwedge_{k\in[w]}\left(\left|x_k - \frac{\sum \mathbf{x}}{w}\right| \leq \frac{1}{2}\lg\lg w + 3 - \rho\right)\right) \bigwedge \left(\bigwedge_{\pi\in\mathcal{P}}\left(\bigvee_{\mathsf{b}\in\pi}\left|\overline{x}(\mathsf{b}_r) - \frac{\sum \mathbf{x}}{w}\right| \leq \frac{1}{2}\lg\lg w + 1 - \rho\right)\right)\right]$$
$$\geq \mathbb{P}\left[\bigwedge_{k\in[w]}\left(\left|x_k - \frac{\sum \mathbf{x}}{w}\right| \leq \frac{1}{2}\lg\lg w + 3 - \rho\right)\right] - \mathbb{P}\left[\bigvee_{\pi\in\mathcal{P}}\left(\bigwedge_{\mathsf{b}\in\pi}\left|\overline{x}(\mathsf{b}_r) - \frac{\sum \mathbf{x}}{w}\right| > \frac{1}{2}\lg\lg w + 1 - \rho\right)\right]$$
$$\geq \mathbb{P}\left[\bigwedge_{k\in[w]}\left(\left|x_k - \frac{\sum \mathbf{x}}{w}\right| \leq \frac{1}{2}\lg\lg w + 3 - \rho\right)\right] - \frac{8}{w}. \qquad \square$$

For each $1 \leq \rho \leq \frac{\lg\lg w}{2} - 6$, let $\mathbf{x}(\rho), \mathbf{y}(\rho)$ be the input and output vector of $\mathcal{L}_\rho$, respectively. We shall prove by induction that for every ρ, with $1 \leq \rho \leq \frac{\lg\lg w}{2} - 6$,

$$\mathbb{P}\left[\bigwedge_{k\in[w]}\left(\left|y_k(\rho) - \frac{\sum \mathbf{x}}{w}\right| \leq \frac{1}{2}\lg\lg w + 2 - \rho\right)\right] \geq 1 - \frac{8\rho+1}{w}.$$

For the basis case ($\rho = 1$), by Lemmas 4.5 and 6.9,

$$\mathbb{P}\left[\bigwedge_{k\in[w]}\left(\left|y_k(1) - \frac{\sum \mathbf{x}}{w}\right| \leq \frac{1}{2}\lg\lg w + 1\right)\right]$$
$$\geq \mathbb{P}\left[\bigwedge_{k\in[w]}\left(\left|x_k(1) - \frac{\sum \mathbf{x}}{w}\right| \leq \frac{1}{2}\lg\lg w + 2\right)\right] - \frac{8}{w}$$
$$\geq 1 - \frac{9}{w}.$$

Assume inductively that the claim holds for $\rho - 1$. For the induction step, by Lemma 6.9 and the induction hypothesis,

$$\mathbb{P}\left[\bigwedge_{k\in[w]}\left(\left|y_k(\rho) - \frac{\sum \mathbf{x}}{w}\right| \leq \frac{1}{2}\lg\lg w + 2 - \rho\right)\right]$$
$$\geq \mathbb{P}\left[\bigwedge_{k\in[w]}\left(\left|x_k(\rho) - \frac{\sum \mathbf{x}}{w}\right| \leq \frac{1}{2}\lg\lg w + 2 - \rho + 1\right)\right] - \frac{8}{w}$$
$$\geq \left(1 - \frac{8(\rho-1)+1}{w}\right) - \frac{8}{w} = 1 - \frac{8\rho+1}{w},$$

and the induction is complete. This implies that for $\rho = \frac{1}{2}\lg\lg w - 6$, $\mathbb{P}\left[\bigwedge_{k\in[w]}\left(\left|y_k\left(\frac{\lg\lg w}{2}-6\right)-\frac{\sum \mathbf{x}}{w}\right|\leq 8\right)\right]\geq 1-\frac{8(\frac{\lg\lg w}{2}-6)+1}{w}$; hence, the output of the second CCC_w is 16-smooth with probability at least $1-\frac{4\lg\lg w-47}{w}$. □

7. IMPROBABILITY OF 1-SMOOTHING

We now use elementary arguments to show:

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

PROOF. Fix a randomized network B_w. Choose two distinct integers $0 \leq i, j \leq w-1$ uniformly at random. Define

$$\mathbf{x}_{i,j} = \langle 1,\ldots,1,\underbrace{0}_{\text{component } i},1,\ldots,1,\underbrace{2}_{\text{component } j},1,\ldots,1\rangle,$$

so, the input vector $\mathbf{x}_{i,j}$ is a random variable. For each layer $\ell, 1 \leq \ell \leq \mathsf{d}(\mathsf{B}_w)$, denote by $\mathcal{E}_\ell$ the event that there is a balancer b in layer ℓ whose inputs are 0 and 2. Clearly, $\mathsf{B}_w(\mathbf{x}_{i,j})$ is 1-smooth if and only if there is a layer ℓ such that $\mathcal{E}_\ell$ occurs. By the Law of Conditional Alternatives,

$$\begin{aligned}&\mathbb{P}\left[\mathsf{B}_w(\mathbf{x}) \text{ is 1-smooth}\right]\\ =&\sum_{\substack{0\leq\hat{\imath},\hat{\jmath}\leq w-1\\ \hat{\imath}\neq\hat{\jmath}}}\mathbb{P}\left[\mathbf{x}=\mathbf{x}_{\hat{\imath},\hat{\jmath}}\right]\cdot\mathbb{P}\left[\mathsf{B}_w(\mathbf{x}) \text{ 1-smooth} \mid \mathbf{x}=\mathbf{x}_{\hat{\imath},\hat{\jmath}}\right]\\ =&\sum_{\substack{0\leq\hat{\imath},\hat{\jmath}\leq w-1\\ \hat{\imath}\neq\hat{\jmath}}}\frac{1}{w(w-1)}\cdot\mathbb{P}\left[\mathsf{B}_w(\mathbf{x}) \text{ 1-smooth} \mid \mathbf{x}=\mathbf{x}_{\hat{\imath},\hat{\jmath}}\right].\end{aligned}$$

LEMMA 7.2. *For each layer* $\ell \geq 1$, $\mathbb{P}\left[\mathcal{E}_\ell\right] \leq \frac{1}{w-1}$.

PROOF. We first prove:

LEMMA 7.3. *Consider a pair* $i, j \in [w], i \neq j$. *Then, for each layer* ℓ, $1 \leq \ell \leq \mathsf{d}(\mathsf{B}_w)$, $\mathbb{P}\left[\mathbf{x}(\ell) = \mathbf{x}_{i,j}\right] \leq \frac{1}{w(w-1)}$.

By the Union Bound and Lemma 7.3,

$$\begin{aligned}&\mathbb{P}\left[\mathcal{E}_\ell\right]=\mathbb{P}\left[\bigvee_{\mathsf{b}\in\ell}\left(\{x_1(\mathsf{b}),x_2(\mathsf{b})\}=\{0,2\}\right)\right]\\ \leq&\sum_{\mathsf{b}\in\ell}\mathbb{P}\left[(x_1(\mathsf{b})=0\wedge x_2(\mathsf{b})=2)\vee(x_1(\mathsf{b})=2\wedge x_2(\mathsf{b})=0)\right]\\ \leq&\sum_{\mathsf{b}\in\ell}\left(\frac{1}{w(w-1)}+\frac{1}{w(w-1)}\right)\ \leq\ \frac{1}{w-1}.\end{aligned}$$

□

Clearly, by the Union Bound and Lemma 7.2,

$$\mathbb{P}\left[\mathsf{B}_w(\mathbf{x}) \text{ is 1-smooth }\right]\leq\sum_{\ell=1}^{\mathsf{d}(\mathsf{B}_w)}\mathbb{P}\left[\mathcal{E}_\ell\right]\leq\mathsf{d}(\mathsf{B}_w)\cdot\frac{1}{w-1}.$$

Hence, there is a pair $0 \leq \hat{\imath}, \hat{\jmath} \leq w-1, \hat{\imath} \neq \hat{\jmath}$ such that $\mathbb{P}\left[\mathsf{B}_w(x) \text{ is 1-smooth} \mid x = x_{\hat{\imath},\hat{\jmath}}\right] \leq \frac{\mathsf{d}(\mathsf{B}_w)}{w-1}$. □

8. EPILOGUE

We presented a thorough study of the impact of randomization in smoothing networks. We proved a tight (up to a small additive constant) bound of $\lg\lg w + \Theta(1)$ on the smoothness of the popular block network. As our main result, we established an upper bound of 16 on the smoothness of the cascade of two block networks. Finally, we proved that it is impossible to obtain a 1-smoothing randomized network of low depth and sufficiently large probability. Our results reveal the full power of randomization in smoothing networks: randomization can be employed in a practical network to yield a constant upper bound on smoothness.

Our work leaves open a number of interesting questions. On the most concrete level, it would be extremely interesting to establish our conjecture that the cascade of a small number of block networks may result to a 2-smoothing network (with high probability).

Acknowledgments. We are grateful to Costas Busch and the anonymous *PODC 2008* reviewers for helpful comments.

9. REFERENCES

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

[2] M. Ajtai, J. Komlós and E. Szemerédi, "An $O(n \lg n)$ Sorting Network," *Combinatorica,* Vol. 3, pp. 1–19, 1983.

[3] J. Aspnes, M. Herlihy and N. Shavit, "Counting Networks," *Journal of the ACM,* Vol. 41, pp. 1020–1048, 1994.

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

[5] C. Busch and M. Mavronicolas, "An Efficient Counting Network," *Proceedings of the 1st Merged Symposium IPPS/SPDP 1998* pp. 380–384, 1998.

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

[7] M. Dowd, Y. Perl, L. Rudoplh, and M. Saks, "The Periodic Balanced Sorting Network," *Journal of the ACM,* Vol. 36, pp. 738–757, 1989.

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

[9] M. Herlihy and S. Tirthapura, "Randomized Smoothing Networks," *Journal of Parallel and Distributed Computing,* Vol. 66, pp. 626–632, 2006.

[10] M. Herlihy and S. Tirthapura, "Self-Stabilizing Smoothing and Counting Networks," *Distributed Computing,* Vol. 18, pp. 345–357, 2006.

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

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

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

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

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