Electronic Colloquium on Computational Complexity, Report No. 48 (2023)

# A $d^{1/2+o(1)}$ Monotonicity Tester for Boolean Functions on $d$-Dimensional Hypergrids

Hadley Black
University of California, Los Angeles
hablack@cs.ucla.edu

Deeparnab Chakrabarty*
Dartmouth
deeparnab@dartmouth.edu

C. Seshadhri†
University of California, Santa Cruz
sesh@ucsc.edu

April 3, 2023

**Abstract**

Monotonicity testing of Boolean functions on the hypergrid, $f : [n]^d \to \{0, 1\}$, is a classic topic in property testing. Determining the non-adaptive complexity of this problem is an important open question. For arbitrary $n$, [Black-Chakrabarty-Seshadhri, SODA 2020] describe a tester with query complexity $\widetilde{O}(\varepsilon^{-4/3}d^{5/6})$. This complexity is independent of $n$, but has a suboptimal dependence on $d$. Recently, [Braverman-Khot-Kindler-Minzer, ITCS 2023] and [Black-Chakrabarty-Seshadhri, STOC 2023] describe $\widetilde{O}(\varepsilon^{-2}n^3\sqrt{d})$ and $\widetilde{O}(\varepsilon^{-2}n\sqrt{d})$-query testers, respectively. These testers have an almost optimal dependence on $d$, but a suboptimal polynomial dependence on $n$.

In this paper, we describe a non-adaptive, one-sided monotonicity tester with query complexity $O(\varepsilon^{-2}d^{1/2+o(1)})$, *independent* of $n$. Up to the $d^{o(1)}$-factors, our result resolves the non-adaptive complexity of monotonicity testing for Boolean functions on hypergrids. The independence of $n$ yields a non-adaptive, one-sided $O(\varepsilon^{-2}d^{1/2+o(1)})$-query monotonicity tester for Boolean functions $f : \mathbb{R}^d \to \{0, 1\}$ associated with an arbitrary product measure.

*Supported by NSF-CAREER award 2041920
†Supported by NSF DMS-2023495, CCF-1740850, 1839317, 1813165, 1908384, 1909790

ISSN 1433-8092

# Contents

# 1 Introduction

Since its introduction more than two decades ago, the problem of monotonicity testing has attracted an immense amount of attention (see §1.3). In this paper, we focus on the question of testing of Boolean functions $f : [n]^d \to \{0,1\}$ over the $d$-dimensional hypergrid. Here $[n]$ denotes the set $\{1, 2, \ldots, n\}$. Each element $\mathbf{x} \in [n]^d$ is represented as a $d$-dimensional vector with $\mathbf{x}_i \in [n]$ denoting the $i$th coordinate. The partial order of the hypergrid is defined as: $\mathbf{x} \preceq \mathbf{y}$ iff $\mathbf{x}_i \leq \mathbf{y}_i$ for all $i \in [d]$. When $n = 2$, the hypergrid $[n]^d$ is isomorphic to the hypercube $\{0,1\}^d$. A Boolean hypergrid function $f : [n]^d \to \{0,1\}$ is monotone if $f(\mathbf{x}) \leq f(\mathbf{y})$ whenever $\mathbf{x} \preceq \mathbf{y}$.

The distance between two functions $f$ and $g$, denoted $\Delta(f,g)$, is the fraction of points where they differ. A function $f : [n]^d \to \{0,1\}$ is called $\varepsilon$-far from monotone if $\Delta(f,g) \geq \varepsilon$ for all monotone functions $g : [n]^d \to \{0,1\}$. Given a proximity parameter $\varepsilon$ and query access to a function, a *monotonicity tester* accepts a monotone function and rejects a function that is $\varepsilon$-far from monotone. Both should occur with probabilty $\geq 2/3$. If the tester accepts monotone functions with probability 1, it is called *one-sided*. If the tester decides its queries without seeing any responses, it is called *non-adaptive*.

An outstanding open question in property testing is to determine the optimal non-adaptive query complexity of monotonicity testing for Boolean hypergrid functions. While we leave the details of the road to the state of the art to §1.3, here we mention the current best bounds. Black, Chakrabarty, and Seshadhri [BCS18, BCS20] give a $\widetilde{O}(\varepsilon^{-4/3}d^{5/6})$-query tester. Note that the query complexity is independent of $n$. Building on seminal work of Khot, Minzer, and Safra [KMS18], Braverman, Khot, Kindler, and Minzer [BKKM23] and Black, Chakrabarty, and Seshadhri [BCS23] recently give $\widetilde{O}(\varepsilon^{-2}n^3\sqrt{d})$ and $\widetilde{O}(\varepsilon^{-2}n\sqrt{d})$ testers, respectively. Chen, Waingarten, and Xie [CWX17] give an $\widetilde{\Omega}(\sqrt{d})$ lower bound for Boolean monotonicity testing on hypercubes ($n = 2$). Hence, these last bounds are nearly optimal in $d$, but are sub-optimal in $n$. Can one achieve the optimal $\sqrt{d}$ dependence while being independent of $n$?

We answer in the affirmative, giving a non-adaptive, one-sided monotonicity tester for Boolean functions over hypergrids with almost optimal query complexity.

**Theorem 1.1.** *Consider Boolean functions over the hypergrid, $f : [n]^d \to \{0,1\}$. There is a one-sided, non-adaptive tester for monotonicity that makes $\varepsilon^{-2}d^{1/2+O(1/\log\log d)}$ queries.*

Query complexities independent of $n$ allow for monotonicity testing over continuous spaces. Let $\mu = \prod_{i=1}^{d} \mu_i$ be an associated product Lebesgue measure over $\mathbb{R}^d$. A function $f : \mathbb{R}^d \to \{0,1\}$ is measurable if the set $f^{-1}(1)$ is Lebesgue-measurable with respect to $\mu$. The $\mu$-distance of $f$ to monotonicity is defined as $\inf_{g \in \mathcal{M}} \mu(\Delta(f,g))$, where $\mathcal{M}$ is the family of measurable monotone functions and $\Delta$ is the symmetric difference operator. (Refer to Sec. 6 of [BCS20] for more details.) Domain reduction results [BCS20, HY22] show that monotonicity testing over general hypergrids and continuous (measurable) spaces can be reduced to the case where $n = \mathrm{poly}(\varepsilon^{-1}d)$ via sampling. A direct consequence of Theorem 1.1 is the following theorem for continuous monotonicity testing.

**Theorem 1.2.** *Consider Boolean functions $f : \mathbb{R}^d \to \{0,1\}$, with an associated measure $\mu$. There is a one-sided, non-adaptive tester for monotonicity that makes $\varepsilon^{-2}d^{1/2+O(1/\log\log d)}$ queries.*

## 1.1 Path Testers, Directed Isoperimetry, and the Dependence on $n$

All $o(d)$ non-adaptive, one-sided monotonicity testers are *path testers* that check for violations among comparable points which are at a distance from each other. Consider the *fully augmented* directed hypergrid graph defined as follows. Its vertices are $[n]^d$ and its edges connect all pairs $\mathbf{x} \prec \mathbf{y}$ that differ in exactly one coordinate. A path tester picks a random point $\mathbf{x}$ in $[n]^d$, performs a random walk in this directed graph to get another point $\mathbf{y} \succ \mathbf{x}$, and rejects if $f(\mathbf{x}) > f(\mathbf{y})$. The whole game is to lower bound the probability that $f(\mathbf{x}) > f(\mathbf{y})$ when $f$ is $\varepsilon$-far from being monotone. Unlike random walks on undirected graphs, these directed random walks are ill-behaved. In particular, one cannot walk for "too long" and the length of the walk has to be carefully chosen. The approach to analyzing such path testers has two distinct parts.

• *Directed Isoperimetry.* A Boolean isoperimetric theorem relates the volume of a subset of the hypercube/grid, in our case the preimage $f^{-1}(1)$, to the edge and vertex expansion properties of this set in the graph. A directed analogue replaces the volume with the distance to monotonicity, and deals with directed expansion properties. This connection between monotonicity testing and directed isoperimetric theorems was first made explicit in [CS14], which also gave the first $o(d)$ tester on hypercubes.

• *Random walk analysis.* The second part is to use the directed isoperimetric theorem to lower bound the success probability of the path tester. The analogy is: if the (directed) expansion of a set is large, then the probability of a directed random walk starting from a 1 and ending at a 0 is also large. This analysis is subtle and proceeds via special combinatorial substructures in the graph of violations.

The seminal result of Khot, Minzer, and Safra [KMS18] (henceforth KMS) gave near optimal analyses for both parts, for the *hypercube* domain. For the first part, they prove a directed, robust version of the Talagrand isoperimetric theorem. For this section it is not crucial to know this theorem. Rather, what is important is that KMS use this directed isoperimetric theorem to construct "good subgraphs" of the fully augmented hypergrid comprising of violated edges. For the second part mentioned above, KMS relate the success probability of the directed random walk to properties of this subgraph. We shortly give details on this second part.

Coming to hypergrids, one needs to generalize both parts of the analysis, and this offers many challenges. For the first part, Black, Chakrabarty, and Seshadhri [BCS23] generalize the directed Talagrand inequality to the hypergrid domain. Unfortunately, even with this stronger directed Talagrand/isoperimetric bound for hypergrids, the generalization of the KMS random walk analysis only yields a $1/(n\sqrt{d})$ lower bound on the escape probability.

> *The main technical contribution of this paper is a new random walk analysis whose success probability is at least $\varepsilon^2 d^{-(1/2+o(1))}$.*

In what follows, we describe the KMS random walk analysis, the generalization to hypergrids, and the challenges in removing the dependence on $n$. In §1.2 we describe our main technical ideas required to bypass these challenges and obtain the independence from $n$.

**The KMS random walk analysis on $\{0,1\}^d$ in a nutshell.** For simplicity, let's assume $\varepsilon$ is a small constant so that we ignore the dependence on $\varepsilon$. Using the directed isoperimetric theorem, KMS extract a large "good subgraph" of violations. A violation subgraph $G = (\boldsymbol{X}, \boldsymbol{Y}, E)$ is a bipartite graph where $\forall \mathbf{x} \in \boldsymbol{X}, f(\mathbf{x}) = 1$, $\forall \mathbf{y} \in \boldsymbol{Y}$, $f(\mathbf{y}) = 0$, and all edges in $E$ are hypercube edges. A good subgraph is a violation subgraph that satisfies certain lower bounds on the total number of edges and has an approximate regularity property. The specifics are a bit involved (Definition 6.1-6.3, in [KMS18]), but it is most instructive to think of the simplest good subgraph. This is a *matching* between $\boldsymbol{X}$ and $\boldsymbol{Y}$ where $|\boldsymbol{X}| = |\boldsymbol{Y}| = \Omega(2^d)$.

When the good subgraph is a matching, KMS show that a random walk of length $\tau = \widetilde{\Theta}(\sqrt{d})$ succeeds in finding a violation with $\widetilde{\Omega}(d^{-1/2})$ probability. A key insight in the analysis is the notion of $\tau$-*persistence*: a vertex $\mathbf{x}$ is $\tau$-persistent if a $\tau$-length directed random walk leads to a point $\mathbf{z}$ where $f(\mathbf{x}) = f(\mathbf{z})$ with constant probability. Using a simple argument based on the influence of the function, KMS argue that an average directed random walk has $\lessapprox \tau/\sqrt{d} = o(1)$ influential edges. Using Markov's inequality, at most $o(2^d)$ points in $\{0,1\}^d$ can be non-persistent. Let us remove all non-persistent points and their matched partners from $\boldsymbol{X}$ and $\boldsymbol{Y}$, to get a matching between $\boldsymbol{X}'$ and $\boldsymbol{Y}'$. All points in $\boldsymbol{X}'$ and $\boldsymbol{Y}'$ are persistent, and given the size bounds, $|\boldsymbol{X}'| = |\boldsymbol{Y}'|$ is still $\Omega(2^d)$.

With $\Omega(1)$ probability, the tester starts from $\mathbf{x} \in \boldsymbol{X}'$. Note that $f(\mathbf{x}) = 1$. Let the matched partner of $\mathbf{x}$ be $\mathbf{y}$. Let $i$ be the dimension of the violated edge $(\mathbf{x}, \mathbf{y})$. With probability roughly $\tau/d = \widetilde{\Omega}(d^{-1/2})$, the directed walk will cross the $i$th dimension. Let us condition on this event. We can interpret the random walk as traversing the edge $(\mathbf{x}, \mathbf{y})$, and then taking a $(\tau - 1)$-length directed walk from $\mathbf{y}$ to reach the destination $\mathbf{y}'$. (Note that we do not care about the specific order of edges traversed by the random walk. We only care about the value at the destination.) Since $\mathbf{y}$ is $\tau$-persistent[1], with $\Omega(1)$ probability the final destination $\mathbf{y}'$ will satisfy $f(\mathbf{y}') = f(\mathbf{y}) = 0$. Putting it all together, the tester succeeds with probability $\widetilde{\Omega}(d^{-1/2})$.

We stress that the above analysis discards any non-persistent vertex in the original matching. Hence, it is critical that the good subgraph is a sufficiently large matching. In general, a good subgraph might not be a matching. However, the KMS directed isoperimetric theorem can be used to obtain a subgraph of violated edges with the following property: if the maximum degree is $\Delta$, then the number of edges is $\Omega(\sqrt{\Delta}2^d)$; when $\Delta = 1$, the subgraph is indeed a large matching. One can then argue that the random walk of length[2] $\tau \approx \sqrt{d/\Delta}$ has success probability is $\widetilde{\Omega}(d^{-1/2})$. In all cases, the analysis needs an interplay between the notion of persistence, the size of the sets $\boldsymbol{X}, \boldsymbol{Y}$, and the degrees in the good subgraph.

**The challenge in hypergrids.** As mentioned earlier, [BCS23] proves an isoperimetric theorem for hypergrids generalizing the one in [KMS18]. Using similar techniques to the hypercube case, one can construct "good subgraphs" of the fully augmented hypergrid. The definition is involved (Theorem 7.8 in [BCS23]), but the simplest case is again a violation matching of $(\boldsymbol{X}, \boldsymbol{Y}, E)$ of size $|\boldsymbol{X}| = |\boldsymbol{Y}| = \Omega(n^d)$. Note that the matched pairs $(\mathbf{x}, \mathbf{y})$ are axis-aligned, that is, differ in exactly one coordinate $i$. But $\mathbf{y}_i - \mathbf{x}_i$ is an integer in $\{1, 2 \ldots, n-1\}$.

In the hypergrid, the directed random walk must necessarily perform "jumps". At each step, the walk changes a chosen coordinate to a random larger value. One can generalize the hypercube persistence arguments to show that with constant probability, a $\tau = \widetilde{\Theta}(\sqrt{d})$-step random walk will result in both endpoints having the same value. And so, like before, we can remove all "non-persistent" points to end up with an $\Omega(n^d)$ violation matching $(\boldsymbol{X}', \boldsymbol{Y}')$ where all vertices are $\tau$-persistent.

The tester picks $\mathbf{x} \in \boldsymbol{X}'$ with $\Omega(1)$ probability. Let $\mathbf{y}$ be its matched partner, which differs in the $i$th coordinate. If the number of steps is $\tau$, then with $\tau/d \approx \widetilde{\Omega}(d^{-1/2})$ probability, the walk will choose to move along the $i$th coordinate. Conditioned on this event, we would like to relate the random walk to persistent walk from $\mathbf{y}$. However, there is only a $1/n$ chance that the *length* jumped along that coordinate will be the jump $\mathbf{y}_i - \mathbf{x}_i$. One loses an extra $n$ factor in the success probability, and indeed, this is the high-level analysis of the $\widetilde{O}_\varepsilon(n\sqrt{d})$-tester from [BCS23] (at least for the case of the matching).

How does one get rid of this dependence on $n$? At some level, there is no (simple) way around this impasse. If $\mathbf{y}_i - \mathbf{x}_i$ is, say $\Theta(n)$, we cannot relate the walk from $\mathbf{x}$ to a (persistent) walk from $\mathbf{y}$ without losing this $n$ factor. If one desires to be free of the parameter $n$, then one needs to consider the *internal points* in the segment $(\mathbf{x}, \mathbf{y})$. But all internal points could be non-persistent. Even though most internal points $\mathbf{z}$ in the segment $(\mathbf{x}, \mathbf{y})$ may be 0-valued, a $(\tau - 1)$-step walk from $\mathbf{z}$ could lead to 1-valued points. So the final pair won't be a violation.

[1]Technically, one needs $(\tau - 1)$-persistence, which holds from $\tau$-persistence.

[2]The algorithm tries all $O(\log d)$ walks of lengths which are a power of 2

One may think that since the matching size was large ($\approx n^d$), perhaps the "interior" (the union of the interiors of the matching segments) would also be large and most of the internal nodes would be pesistent. Unfortunately, that may not be the case, and the following is an illustrative example. We define a Boolean hypergrid function $f$ and an associated violation matching iteratively. Let $n \leq d/\ln d$. Start with all function values undefined. If $\mathbf{x}_1 = 1$, set $f(\mathbf{x}) = 1$. If $\mathbf{x}_1 = n$, set $f(\mathbf{x}) = 0$. Take the natural violation matching between these points. For every undefined point $\mathbf{x}$: if $\mathbf{x}_2 = 1$, set $f(\mathbf{x}) = 1$ and if $\mathbf{x}_n = 1$, set $f(\mathbf{x}) = 0$. Iterating over all coordinates, we define the function at all points "on the surface". In the "interior", where $\forall i$, $\mathbf{x}_i \notin \{1, n\}$, we set $f$ arbitrarily. The interior has size $n^d \cdot (1 - 2/n)^d \approx n^d \exp(-2 \ln d) \leq n^d/d^2$. This is a tiny fraction of the domain, while the matching has size $\Omega(n^d)$. Hence, it is possible to have a large violation matching such that the union of (strict) interiors is vanishingly small.

In sum, to get rid of the dependence on $n$ requires a new set of ideas.

## 1.2 Our Main Ideas

As in §1.1, we begin with the basic case of a violation matching $G = (\boldsymbol{X}, \boldsymbol{Y}, E)$ of size $\Omega(n^d)$ in the fully augmented hypergrid. The general case will be discussed at the end of this section. We set $\tau = \widetilde{\Theta}(\sqrt{d})$. Using the persistence and Markov inequality arguments, we can assume that all points in $\boldsymbol{X} \cup \boldsymbol{Y}$ are $\tau$-persistent.

Let us describe the tester (Algorithm 1). A $\tau$-step/length upwalk from a point $\mathbf{x} \in [n]^d$ first chooses $\tau$ coordinates at random to increase. To pick the increment on each coordinate, we apply a standard technique for hypergrid property testing. For each coordinate indepedently, the tester picks a random power of 2, and then picks a uniform random increment less than the random power of 2. Analogously, the tester performs downwalks. Unlike the path testers for hypercubes, it will be important that our tester performs both upwalk and downwalks. In our analysis, we will relate the success probabilities of these different walks.

The tester also performs *shifted path tests*. First, it finds a pair $(\mathbf{x}, \mathbf{w})$ using the directed random walk. Then, it samples a random shift vector $\mathbf{s} \in [n]^d$. This is a random vector with $\tau$ non-zero coordinates. The shifted pair is $(\mathbf{x} - \mathbf{s}, \mathbf{w} - \mathbf{s})$. Note that the shifted pair is also comparable, and is equivalent to generating points by correlated random downwalks from $\mathbf{x}$ and $\mathbf{w}$.

**Mostly-zero-below points, and red edges.** The following is a key definition: we call a point $\mathbf{w}$ *mostly-zero-below* for length $\tau$, or simply $\tau$-mzb, if a $\tau$-length downwalk from $\mathbf{w}$ leads to a zero with $\geq 0.9$ probability (Definition 4.1). Suppose an upwalk of length $\tau$ from a point $\mathbf{x} \in \boldsymbol{X}$ reaches an $\tau$-mzb point $\mathbf{w}$. Then, a random shift $(\mathbf{x} - \mathbf{s}, \mathbf{w} - \mathbf{s})$ has a constant probability of being a violation. The reason is (i) $\Pr[f(\mathbf{x} - \mathbf{s}) = f(\mathbf{x}) = 1] \geq 0.9$ because $\mathbf{x}$ is $\tau$-persistent, and (ii) $\Pr[f(\mathbf{w} - \mathbf{s}) = 0] \geq 0.9$ because $\mathbf{w}$ is $\tau$-mostly-zero-below. By a union bound, the tester will find a violation with constant probability (conditioned on discovering the pair $(\mathbf{x}, \mathbf{w})$).

To formalize this analysis, we define a matching edge $(\mathbf{x}, \mathbf{y})$ to be *red* if it satisfies the following condition. For a constant fraction of the interior points $\mathbf{z}$ in the segment $(\mathbf{x}, \mathbf{y})$, a $(\tau - 1)$-length upwalk ends at a $\tau$-mzb point with constant probability (Definition 4.2). If there are $\Omega(n^d)$ red matching edges, we can argue that the tester succeeds with the desired probability. Firstly, with probability $\Omega(1)$, the tester starts the walk at and endpoint $\mathbf{x}$ of a red edge. Let the matched edge be $(\mathbf{x}, \mathbf{y})$. With probability $\tau/d \approx d^{-1/2}$, the walk will cross the dimension corresponding to $(\mathbf{x}, \mathbf{y})$. Conditioned on this event, we can interpret the walk as first moving to a random interior point $\mathbf{z}$ in the segment $(\mathbf{x}, \mathbf{y})$ and then taking a $(\tau - 1)$-length upwalk from $\mathbf{z}$ to get to the point $\mathbf{z}'$. (Refer to Fig. 1.) Since the edge was red, with constant probability, $\mathbf{z}'$ is $\tau$-mzb. Consider a random shift of $(\mathbf{x}, \mathbf{z}')$, shown as $(\mathbf{x} - \mathbf{t}, \mathbf{z}' - \mathbf{t})$ in Fig. 1. As discussed in the previous paragraph, this shifted pair is a violation with constant probability. All in all, the tester succeeds with $\Omega(d^{-1/2})$ probability.

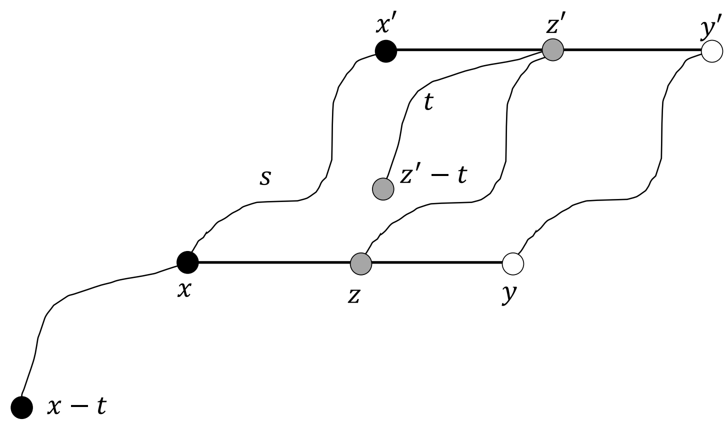


Figure 1: *This figure shows the key argument that either upwalks + downshifts, or downwalks find violations. The edge $(\mathbf{x}, \mathbf{y})$ is in the initial violation matching. Parallel curves of the same shape denote the same shift. So $\mathbf{x}' = \mathbf{x} + \mathbf{s}$, $\mathbf{y}' = \mathbf{y} + \mathbf{s}$, and $\mathbf{z}' = \mathbf{z} + \mathbf{s}$. Similarly, we see both $\mathbf{x}$ and $\mathbf{z}'$ shifted below by $\mathbf{t}$. The 1-valued points are colored black and the 0-valued points are colored white. Gray points do not have an a priori guarantee on function value. If $\mathbf{z}'$ is* mzb*, then $f(\mathbf{z}' - \mathbf{t}) = 0$ with high probability. In this case, $(\mathbf{x} - \mathbf{t}, \mathbf{z}' - \mathbf{t})$ is a likely violation. If not, then $(\mathbf{z}' - \mathbf{t}, \mathbf{y}')$ is a likely violation.*

But what if there are no red edges? This takes us to the next key idea of our paper: *translations* of violation subgraphs.

**Translations of violation subgraphs, and blue edges.** Suppose most of the matching edges edges $(\mathbf{x}, \mathbf{y})$ are not red. So, for most points $\mathbf{z}$ in the segment $(\mathbf{x}, \mathbf{y})$, a $(\tau - 1)$-length walk does not reach a $\tau$-mzb point. Fix one such walk, which can be described by an "up-shift" $\mathbf{s}$. So the walk from $\mathbf{z}$ reaches $\mathbf{z}' := \mathbf{z} + \mathbf{s}$.

Consider the corresponding shift of the full edge $(\mathbf{x}, \mathbf{y})$ to $(\mathbf{x}', \mathbf{y}')$, where $\mathbf{x}' = \mathbf{x} + \mathbf{s}$ and $\mathbf{y}' = \mathbf{y} + \mathbf{s}$. Refer to Fig. 1. What can we say about this edge? Since both $\mathbf{x}$ and $\mathbf{y}$ are up-persistent, with high probability both $f(\mathbf{x}') = f(\mathbf{x}) = 1$ and $f(\mathbf{y}') = f(\mathbf{y}) = 0$. Observe that most internal points $\mathbf{z}'$ in $(\mathbf{x}', \mathbf{y}')$ are not mostly-zero-below. Consider a $\tau$-length downward walk from $\mathbf{z}'$, whose destination can be represented as $\mathbf{z}' - \mathbf{t}$ (for a downshift $\mathbf{t}$). With probability $\geq 0.1$, $f(\mathbf{z}' - \mathbf{t}) = 1$.

Recall, the tester performs a downward random walk (Algorithm 1, Step 3). Suppose this walk starts at $\mathbf{y}'$. With probability $\approx \tau/d \approx d^{-1/2}$, the walk moves (downward) in the $i$th coordinate with constant probability. Conditioned on this, the walk ends up at a point $\mathbf{z}' - \mathbf{t}$. As discussed above, $\mathbf{z}'$ is likely to be not mostly-zero-below. Hence $f(\mathbf{z}' - \mathbf{t}) = 1$ with constant probability, and the tester discovers the violating pair $(\mathbf{z}' - \mathbf{t}, \mathbf{y}')$.

Fig. 1 summarizes the above observations. If $(\mathbf{x}, \mathbf{y})$ is red, then the pair $(\mathbf{x} - \mathbf{t}, \mathbf{z}' - \mathbf{t})$ is likely to a violation. If $(\mathbf{x}, \mathbf{y})$ is not red, then the pair $(\mathbf{z}' - \mathbf{t}, \mathbf{y}')$ is a likely violation. This motivates the definition of our *blue* edges. We call a violating edge blue, if for a constant fraction of points in the interior, a downward walk leads to a 1-point with constant probability (Definition 4.3). We argued above that if the edge $(\mathbf{x}, \mathbf{y})$ in the violation matching was not red, then a random shift or translation up to $(\mathbf{x}', \mathbf{y}')$ leads to a blue edge. If most edges in our original violation matching were not red, then we could translate "all these edges together" to get a (potentially) new large violation subgraph. If most of these new edges are blue, then the downward walk would catch a violation with $\approx d^{-1/2}$ probability.

What does it mean to translate "all edges together"? In particular, how do we pin down this new violation matching? We use ideas from network flows. Through the random translation, every non-red edge $(\mathbf{x}, \mathbf{y})$ in the original violation matching leads to a *distribution* over blue edges $(\mathbf{x}', \mathbf{y}')$. We treat this as a fractional flow on these blue edges. If the original matching had few red edges, we can construct a large collection of blue edges sustaining a large flow. Integrality of flow implies there must be another large violation matching in the support of this distribution whose edges are blue. This is the essence of the "red/blue" lemma (Lemma 4.4).

Putting it together, suppose $G = (\boldsymbol{X}, \boldsymbol{Y}, E)$ is a large violation matching. Either the upwalk with a shift or the downwalk succeeds with probability $\approx d^{-1/2}$.

**Lopsided violation subgraphs and translation again.** We have discussed the situation of a large violation matching $G = (\boldsymbol{X}, \boldsymbol{Y}, E)$ with $|\boldsymbol{X}| = |\boldsymbol{Y}| = \Omega(n^d)$. However, such a large matching may not exist. Instead, the directed isoperimetric theorems imply the existence of a "good subgraph" with bounded maximum degree and many edges. These graphs $G = (\boldsymbol{X}, \boldsymbol{Y}, E)$ may be lopsided with $|\boldsymbol{X}| \ll |\boldsymbol{Y}|$. This causes a significant headache for our algorithm, and once again, the issue is persistence. The good subgraph could have $|\boldsymbol{X}| \approx n^d/\sqrt{d}$, $|\boldsymbol{Y}| \approx n^d$, and edges that are structured as follows. All edges incident to an individual $\mathbf{y} \in \boldsymbol{Y}$ are aligned along the same dimension. For the path tester to find a violation starting from any $\mathbf{y} \in \boldsymbol{Y}$, it must take a walk of length $\tau = \widetilde{\Omega}(\sqrt{d})$.

Unlike in [KMS18] or in [BCS23], the tester must run both the upwalk and downwalk. In the situation of Fig. 1, it is critical that both upwalks and downwalks have the same length. In the lopsided good subgraph indicated above, the walk length is $\widetilde{\Omega}(\sqrt{d})$. For this length, the fraction of non-persistent points could be $\widetilde{\Omega}(1)$. In particular, all the vertices in $\boldsymbol{X}$ could be non-persistent with respect to this length. Thus, the upward walk + downward shift is no longer guaranteed to work. (In Fig. 1, we are no longer guaranteed that $f(\mathbf{x} - \mathbf{t}) = 1$. To ensure that, the walk must be much shorter. But in that case, the walk from $\mathbf{y}'$ is unlikely to cross the $i$th dimension.)

To cross this hurdle, we use the translation idea again. Suppose we had a lopsided violation subgraph $G = (\boldsymbol{X}, \boldsymbol{Y}, E)$ with $|\boldsymbol{X}| \ll |\boldsymbol{Y}|$. For the walk length $\tau$ determined by $\boldsymbol{Y}$, most vertices in $\boldsymbol{X}$ are not down persistent. However, the vertices in $\boldsymbol{X}$ must be up persistent for otherwise the upward walk would succeed (Claim 6.6). Therefore, we can take upward translations of $G$ and again using network flow arguments alluded to in the previous paragraph, we are able to construct another violation subgraph $G' = (\boldsymbol{X}', \boldsymbol{Y}', E')$ that satisfies the following properties. Firstly, $G'$ is "structurally" similar to $G$, in terms of degree bounds and the number of edges. Either vertices in $\boldsymbol{X}'$ are $\tau$-down persistent or $|\boldsymbol{X}'| \geq 2|\boldsymbol{X}|$. We refer to this as the 'persist-or-blow-up' lemma (Lemma 6.7). The argument is somewhat intricate and requires a delicate balance of parameters. An interesting aspect is that we can either beat the usual Markov upper bound for persistent vertices, or improve the parameters of the violation graph. By iterations of the lemma, we can argue the existence of a violation subgraph with all the desired persistent properties. Then, the analysis akin to the matching case generalizes to give the desired result.

**Thresholded degrees, peeling, and the $d^{o(1)}$ loss.** Another gnarly issue with hypergrids is the distinction between degree and "thresholded degree". The relevant "degree" of a vertex $\mathbf{x}$ (for the path tester analysis) in a violation subgraph is not the *number* of edges incident on it, but rather the number of different dimensions $i$ so that there is an $i$-edge incident on it. We refer to this quantity as the "thresholded degree", and it is between 1 and $d$. Note that the standard degree could be as large as $nd$. It is critical one uses thresholded degree for the path tester analysis, to avoid the linear dependence on $n$ in our calculations. Observe that for the matching case, these degrees are identical, making the analysis easier.

While the path tester analysis works with thresholded degree, the flow-based translation arguments alluded to above need to use normal degrees. In particular, we can use flow-arguments to relate the bound the standard degree of the new violation subgraphs. But we cannot a priori do so for the thresholded degree.

To argue about the thresholded degree, we begin with a stronger notion of a good subgraph called the seed regular violation subgraph (Lemma 6.1). This subgraph satisfies specific conditions for both thresholded and standard degrees of the vertices. It is in the construction of the seed graph where we lose the $d^{o(1)}$ factor.

**Roadmap.** Here is the roadmap of the whole analysis. We use the isoperimetric theorem in [BCS23] to prove the existence of the seed regular graph (Lemma 6.1). This subgraph may not have the desired persistence properties, so we apply the persist-or-blow-up lemma, Lemma 6.7, to obtain a more robust graph $G'$. This graph $G'$ may have lots of red edges, in which case it is a "nice red subgraph" (Definition 4.5), and then the upwalk + down-shift (Step 4 in Algorithm 1) succeeds with good probability. Otherwise, we apply the "red/blue" lemma to obtain a "nice blue subgraph" (Definition 4.6), and then the downwalk (Step 3 in Algorithm 1) succeeds with good probability. Of course, the lopsidedness in the seed graph can be $|\boldsymbol{X}| \gg |\boldsymbol{Y}|$ in which case the argument is analogous, except one of Step 2 or Step 5 in Algorithm 1 succeed.

### 1.3 Related Work

Monotonicity testing, and in particular that of Boolean functions on the hypergrid, has been studied extensively in the past 25 years [Ras99, EKK+00, GGL+00, DGL+99, LR01, FLN+02, HK03, AC06, HK08, ACCL07, Fis04, SS08, Bha08, BCSM12, FR10, BBM12, RRS+12, BGJ+12, CS13, CS14, CST14, BRY14a, BRY14b, CDST15, CDJS17, KMS18, BB21, CWX17, BCS18, BCS20, BKR20, HY22, BKKM23, BCS23]. Most of the early works focused on the special case of hypercubes $\{0,1\}^d$. Early works defined the problem and described a $O(d)$ tester [Ras99, GGL+00]. This was improved by [CS14] to give an $\widetilde{O}_\varepsilon(d^{7/8})$ tester and this paper introduced the connection to directed isoperimetry. Subsequently, [KMS18] described their $\widetilde{O}_\varepsilon(\sqrt{d})$ non-adaptive, one-sided tester via the directed robust version of Talagrand's isoperimetric theorem, and this dependence on $d$ is tight even for two-sided testers [FLN+02, CDST15, CWX17]. The best lower bound for adaptive testers is $\Omega(d^{1/3})$ [CWX17, BB21].

Dodis et al [DGL+99] were the first to define the problem of monotonicity testing on general hypergrids, and they gave a non-adaptive, one-sided $O((d/\varepsilon)\log^2(d/\varepsilon))$-query tester for the Boolean range. Thus, it was known from the beginning that independence of $n$ is achievable for Boolean monotonicity testing. Berman, Raskhodnikova, and Yaroslavtsev improved the upper bound to $O((d/\varepsilon)\log(d/\varepsilon))$ [BRY14a]. They also show a non-adaptive lower bound of $\Omega(\log(1/\varepsilon)/\varepsilon)$ and prove an adaptivity gap by giving an adaptive $O(1/\varepsilon)$-query tester for constant $d$.

The first $o(d)$ tester for hypergrids was given by Black, Chakrabarty, Seshadhri [BCS18]. Using a directed Margulis inequality, they achieve a $\widetilde{O}_\varepsilon(d^{5/6}\log n)$ upper bound. In a subsequent result, they introduce the concept of domain reduction and show that $n$ can be reduced to $\mathrm{poly}(d\varepsilon^{-1})$ by subsampling the hypergrid [BCS20]. Harms and Yoshida gave a substantially simpler proof of the domain reduction theorem, though their result is not "black-box" [HY22].

Most relevant to our work are the independent, recent results of Black, Chakrabarty, Seshadhri, and Braverman, Kindler, Khot, Minzer [BCS23, BKKM23]. These results give $\widetilde{O}(\mathrm{poly}(n)\sqrt{d})$ query testers, but with different approaches. The former follows the KMS path, and proves a new directed Talagrand inequality over the hypergrid. This theorem is a key tool in our result. The result of [BKKM23] follows a different approach, via reductions to hypercube monotonicity testing. This is a tricky and intricate construction; naive subsampling approaches to reduce to the hypercube are known to fail (see Sec. 8 of [BCS20]). Instead, their result uses a notion of "monotone" embeddings that embed functions over arbitrary product domains to hypercube functions, while preserving the distance to monotonicity. However, these embeddings increase the dimension by $\mathrm{poly}(n)$, which appears to be inherent.

### 1.4 Discussion

It is an interesting question to see if the $d^{o(1)}$ dependence can be reduced to polylogarithmic in $d$. As mentioned above, the loss arises due to our need for a stronger notion of a "good subgraph". Nevertheless, we feel one could obtain an $\widetilde{O}(\varepsilon^{-2}\sqrt{d})$-tester. In Section 8 of their paper, [BCS23] conjecture a stronger "weighted" isoperimetric theorem which would imply a $\widetilde{O}(\varepsilon^{-2}\sqrt{d})$-tester. Our work currently has no bearing on that conjecture, and that is still open.

At a qualitative level, our work and the result in [BCS23] indicates the Boolean monotonicity testing question on the hypergrid seems more challenging than on the hypercube. Is there a quantitative separation possible? It is likely that non-adaptive monotonicity testing for general hypergrids is harder than hypercubes by "only" a $\log d$ factor. The gap between the non-adaptive upper and lower bounds even for hypercubes is $\mathrm{poly}(\log d)$. So, achieving this separation between hypergrids and hypercubes seems quite challenging, as it would require upper and lower bounds of far higher precision.

## 2 Random Walks and the Monotonicity Tester

Without loss of generality[3] we assume that $n$ is a power of 2. We use $x \in_R S$ to denote choosing a uniform random element $x$ from the set $S$. Abusing notation, we define intervals in $\mathbb{Z}_n$ by wrapping around. So, if $1 \leq i \leq n < j$, then the interval $[i, j]$ in $\mathbb{Z}_n$ is the set $[i, n] \cup [1, (j-1) \pmod{n}]$.

The directed (lazy) random walk distribution in $[n]^d$ that we consider is defined as follows. The distribution induced by this directed walk has multiple equivalent formulations, which are discussed in §3.2.

**Definition 2.1** (Hypergrid Walk Distribution)**.** *For a point $\mathbf{x} \in [n]^d$ and walk length $\tau$, the distribution $\mathcal{U}_\tau(\mathbf{x})$ over $\mathbf{y} \in [n]^d$ reached by an* upward lazy random walk from $\mathbf{x}$ of $\tau$-steps *is defined as follows.*

1. *Pick a uniform random subset $R \subseteq [d]$ of $\tau$ coordinates.*
2. *For each $r \in R$:*
   (a) *Choose $q_r \in_R \{1, 2, \ldots, \log n\}$ uniformly at random.*
   (b) *Choose a uniform random interval $I_r$ in $\mathbb{Z}_n$ of size $2^{q_r}$ such that $\mathbf{x}_r \in I_r$.*
   (c) *Choose a uniform random $c_r \in_R I_r \setminus \{\mathbf{x}_r\}$.*
3. *Generate $\mathbf{y}$ as follows. For every $r \in [d]$, if $r \in R$* and *$c_r > \mathbf{x}_r$, set $\mathbf{y}_r = c_r$. Else, set $\mathbf{y}_r = \mathbf{x}_r$.*

*Analogously, let $\mathcal{D}_\tau(\mathbf{x})$ be the distribution defined precisely as above, but the $>$-sign is replaced by the $<$-sign in step 3. This is the distribution of the endpoint of a* downward lazy random walk from $\mathbf{x}$ of $\tau$-steps.

[3] See Theorem A.1 of [BCS18]. Note this assumption is not crucial, but we choose to use it for the sake of a cleaner presentation.

A crucial step of our algorithm involves performing the exact same random walk, but originating from two different points. We can express our random walk distribution in terms of shifts (rather than destinations) as follows.

**Definition 2.2** (Shift Distributions)**.** *The up-shift distribution from* $\mathbf{x}$*, denoted* $\mathcal{US}_\tau(\mathbf{x})$ *is the distribution of* $\mathbf{x}' - \mathbf{x}$*, where* $\mathbf{x}' \sim \mathcal{U}_\tau(\mathbf{x})$*. The down-shift distribution from* $\mathbf{x}$*, denoted* $\mathcal{DS}_\tau(\mathbf{x})$ *is the distribution of* $\mathbf{x} - \mathbf{x}'$*, where* $\mathbf{x}' \sim \mathcal{D}_\tau(\mathbf{x})$*.*

Note that $\mathcal{U}_\tau(\mathbf{x})$ is equivalent to the distribution of $\mathbf{x} + \mathbf{s}$, where $\mathbf{s} \sim \mathcal{US}_\tau(\mathbf{x})$. Similarly, $\mathcal{D}_\tau(\mathbf{x})$ is equivalent to the distribution of $\mathbf{x} - \mathbf{s}$, where $\mathbf{s} \sim \mathcal{DS}_\tau(\mathbf{x})$. Using Definition 2.1 and Definition 2.2, our tester is defined in Alg. 1.

---

**Algorithm 1** Monotonicity tester for Boolean functions on $[n]^d$

---

**Input:** A Boolean function $f\colon [n]^d \to \{0,1\}$

1. Choose $p \in_R \{0, 1, 2, \ldots, \lceil \log d \rceil\}$ uniformly at random and set $\tau := 2^p$.
2. Run the *upward path test* with walk length $\ell = \tau - 1$ and $\ell = \tau$:
   (a) Choose $\mathbf{x} \in_R [n]^d$ and sample $\mathbf{y}$ from $\mathcal{U}_\ell(\mathbf{x})$.
   (b) If $f(\mathbf{x}) > f(\mathbf{y})$, then reject.
3. Run the *downward path test* with walk length $\ell = \tau - 1$ and $\ell = \tau$:
   (a) Choose $\mathbf{y} \in_R [n]^d$ and sample $\mathbf{x}$ from $\mathcal{D}_\ell(\mathbf{y})$.
   (b) If $f(\mathbf{x}) > f(\mathbf{y})$, then reject.
4. Run the *upward path + downward shift test* with walk length $\ell = \tau - 1$ and $\ell = \tau$:
   (a) Choose $\mathbf{x} \in_R [n]^d$, sample $\mathbf{y}$ from $\mathcal{U}_\ell(\mathbf{x})$, and sample $\mathbf{s}$ from $\mathcal{DS}_{\tau-1}(\mathbf{x})$.
   (b) If $f(\mathbf{x} - \mathbf{s}) > f(\mathbf{y} - \mathbf{s})$, then reject.
5. Run the *downward path + upward shift test* with walk length $\ell = \tau - 1$ and $\ell = \tau$:
   (a) Choose $\mathbf{y} \in_R [n]^d$, sample $\mathbf{x}$ from $\mathcal{D}_\ell(\mathbf{y})$, and sample $\mathbf{s}$ from $\mathcal{US}_{\tau-1}(\mathbf{y})$.
   (b) If $f(\mathbf{x} + \mathbf{s}) > f(\mathbf{y} + \mathbf{s})$, then reject.

---

**Remark 2.3.** *Given a function* $f : [n]^d \to \{0,1\}$*, consider the doubly-flipped function* $g : [n]^d \to \{0,1\}$ *defined as* $g(\mathbf{x}) := 1 - f(\bar{\mathbf{x}})$ *where* $\bar{\mathbf{x}}_i := n - \mathbf{x}_i$*. That is, we swap all the zeros and ones in* $f$*, and then reverse the hypergrid (the all zeros point becomes the all* $n$*'s point and vice-versa). The distance to monotonicity of both* $f$ *and* $g$ *are the same: a pair* $(\mathbf{x}, \mathbf{y})$ *is violating in* $f$ *if and only if* $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ *is violating in* $g$*. In Alg. 1, Step 2 on* $f$ *is the same as Step 3 on* $g$*, and Step 4 on* $f$ *is the same as Step 5 on* $g$*. In our analysis, we will construct a violation subgraph between vertex sets* $\boldsymbol{X}$ *and* $\boldsymbol{Y}$*. Points in* $\boldsymbol{X}$ *are* 1*-valued and points in* $\boldsymbol{Y}$ *are* 0*-valued. If* $|\boldsymbol{X}| \le |\boldsymbol{Y}|$*, then the steps 2, 3, and 4 suffice for the analysis. If* $|\boldsymbol{Y}| \le |\boldsymbol{X}|$*, then (by the same analysis) we run steps 2,3, and 4 on the function* $g$*. This is equivalent to running steps 2, 3, and 5 on the function* $f$*. So, the tester covers both situations, and we can assume wlog that* $|\boldsymbol{X}| \le |\boldsymbol{Y}|$*. This discussion happens in Section 6.1.1.*

Our main result is the following lower bound on the rejection probability of Alg. 1.

**Theorem 2.4** (Main Theorem)**.** *Let $n, \varepsilon^{-1} \leq \text{poly}(d)$. If $f\colon [n]^d \to \{0,1\}$ is $\varepsilon$-far from being monotone, then for any $\delta > (\log\log nd)^{-1}$, Alg. 1 rejects $f$ with probability at least $\varepsilon^2 \cdot d^{-(1/2+O(\delta))}$.*

Theorem 2.4 is proved in §5. We first use Theorem 2.4 to prove our main testing results, Theorem 1.1 and Theorem 1.2.

### 2.1 Proof of Theorem 1.1 and Theorem 1.2

To prove Theorem 1.1, we use the domain reduction Theorem 1.3 of [BCS20], which we state here for ease of reading.

**Theorem 2.5** (Domain Reduction Theorem 1.3, [BCS20])**.** *Suppose $f : [n]^d \to \{0,1\}$ is $\varepsilon$-far from being monotone. Let $k = (\varepsilon^{-1}d)^8$. If $\mathbf{T} = T_1 \times \cdots \times T_d$ is a randomly chosen sub-grid, where for each $i \in [d]$, $T_i$ is a (multi)-set formed by taking $k$ independent, uniform samples from $[n]$, then $\mathbb{E}_{\mathbf{T}}[\varepsilon_{f|_{\mathbf{T}}}] \geq \varepsilon/2$.*

**Remark 2.6.** *We note that [HY22] obtain a more efficient domain reduction result. However, the domain reduction from [BCS20] can be used in a black-box fashion, resulting in a simpler tester.*

*For ease of reading, we give a simplified proof of a weaker version of Theorem 1.1. This proof obtains a tester with an $\varepsilon^{-3}$ dependence, instead of the stated $\varepsilon^{-2}$. A more nuanced argument yields the improved $\varepsilon^{-2}\log(1/\varepsilon)$ dependence, which proves Theorem 1.1 as stated[4]. For details, we refer the reader to Section 7 of [BCS20]. In particular, we run Algorithm 1 in Section 7 of [BCS20] with the sub-routine in line 5 replaced by Alg. 1.*

*Proof. of Theorem 1.1*: Consider the tester which does the following, given $f\colon [n]^d \to \{0,1\}$ and $\varepsilon \in (0,1)$.

1. If $\varepsilon < d^{-1/2}$, then run the $\widetilde{O}(\varepsilon^{-1}d)$ query non-adaptive and 1-sided tester of [DGL+99] or [BRY14a].
2. If $\varepsilon \geq d^{-1/2}$, then set $k = (\varepsilon^{-1}d)^8 \leq d^{12}$ and repeat the following $8\varepsilon^{-1}$ times.
   (a) Sample a $[k]^d$ sub-grid $\mathbf{T} \subseteq [n]^d$ according to the distribution described in Theorem 2.5.
   (b) Run $32 \cdot \varepsilon^{-2} \cdot d^{1/2+O(\delta)}$ iterations of the tester described in Alg. 1 on the restricted function $f|_{\mathbf{T}}$.
3. Accept.

If $\varepsilon < d^{-1/2}$, then the number of queries is $O(\varepsilon^{-1}d) = O(\varepsilon^{-2}d^{1/2})$. We are done in this case.

Assume $\varepsilon \geq d^{-1/2}$. The total number of queries made by this tester is at most $\varepsilon^{-3} \cdot d^{1/2+O(\delta)}$. Clearly, if $f$ is monotone, then the tester will accept, so suppose $\varepsilon_f \geq \varepsilon$. By the domain reduction Theorem 2.5, we have $\mathbb{E}_{\mathbf{T}}[\varepsilon_{f|_{\mathbf{T}}}] \geq \varepsilon/2$. So, $\mathbb{E}_{\mathbf{T}}[1 - \varepsilon_{f|_{\mathbf{T}}}] \leq 1 - \varepsilon/2$ and thus by Markov's inequality,

$$\Pr_{\mathbf{T}}\left[1 - \varepsilon_{f|_{\mathbf{T}}} \geq 1 - \varepsilon/4\right] \leq \frac{1-\varepsilon/2}{1-\varepsilon/4} = \frac{1-\varepsilon/4-\varepsilon/4}{1-\varepsilon/4} \leq 1 - \varepsilon/4.$$

Thus, $\Pr_{\mathbf{T}}[\varepsilon_{f|_{\mathbf{T}}} \geq \varepsilon/4] \geq \varepsilon/4$. Thus, with probability at least $1 - (1-\varepsilon/4)^{8/\varepsilon} \geq 1 - e^{-2}$, some iteration of step (2a) will produce $\mathbf{T}$ such that $\varepsilon_{f|_{\mathbf{T}}} \geq \varepsilon/4$. When this happens, some iteration of step (2b) will reject with probability at least $1 - e^{-2}$, by Theorem 2.4. Thus, the tester rejects $f$ with probability at least $(1-e^{-2})^2 \geq 2/3$. □

The proof of Theorem 1.2 for testing on $\mathbb{R}^d$ follows the exact same argument, using the corresponding domain reduction Theorem 1.4 of [BCS20] for functions over $\mathbb{R}^d$. We omit the proof.

[4] When we invoke Alg. 1, we assume that $\varepsilon \geq d^{-1/2}$ and so $\log 1/\varepsilon = d^{\frac{\log\log 1/\varepsilon}{\log d}} \ll d^{O(1/\log\log d)}$. The factor of $\log 1/\varepsilon$ is absorbed by $d^{O(1/\log\log d)}$ in the query complexity.

# 3 Technical Preliminaries

In this section, we list out preliminary definitions and notations. Throughout the section, we fix a function $f : [n]^d \to \{0, 1\}$ that is $\varepsilon$-far from monotone. For ease of readability, most proofs of this section are in the appendix.

## 3.1 Violation Subgraphs and Isoperimetry

The *fully augmented hypergrid* is a graph whose vertex set is $[n]^d$ where edges connect all pairs that differ in exactly one coordinate. We direct all edges from lower to higher endpoint. The edge $(\mathbf{x}, \mathbf{y})$ is called an *$i$-edge* for $i \in [d]$ if $\mathbf{x}$ and $\mathbf{y}$ differ in the $i$th coordinate. We use $I(\mathbf{x}, \mathbf{y}) = \{\mathbf{z} \colon \mathbf{x} \preceq \mathbf{z} \preceq \mathbf{y}\}$ to denote the points $\mathbf{z}$ in the segment $[\mathbf{x}, \mathbf{y}]$, that is, they are the points which differ from $\mathbf{x}$ and $\mathbf{y}$ only in the $i$th coordinate, and $\mathbf{x}_i \le \mathbf{z}_i \le \mathbf{y}_i$. Given a function $f : [n]^d \to \{0, 1\}$ the edge $(\mathbf{x}, \mathbf{y})$ of the fully augmented hypergrid is a *violating/violated edge* if $f(\mathbf{x}) = 1$ and $f(\mathbf{y}) = 0$.

**Definition 3.1.** *A* violation subgraph *is a subgraph of the fully augmented hypergrid all of whose edges are violations.*

Note that any violation subgraph is a bipartite subgraph, where the bipartition is given by the 1-valued and 0-valued points. We henceforth always express a violation subgraph as $G = (\boldsymbol{X}, \boldsymbol{Y}, E)$ such that $\forall \mathbf{x} \in \boldsymbol{X}$, $f(\mathbf{x}) = 1$ and $\forall \mathbf{y} \in \boldsymbol{Y}$, $f(\mathbf{y}) = 0$. There are a number of relevant parameters of violation subgraphs that play a role in our analysis.

**Definition 3.2.** *Fix a violation subgraph $G = (\boldsymbol{X}, \boldsymbol{Y}, E)$ and a point $\mathbf{x} \in \boldsymbol{X}$.*

- *The* degree *of $\mathbf{x}$ in $G$ is the number of edges in $E$ incident to $\mathbf{x}$ and is denoted as $D_G(\mathbf{x})$.*
- *For any coordinate $i \in [d]$, the $i$-*degree *of $\mathbf{x}$ in $G$ is the total number of $i$-edges in $E$ incident to $\mathbf{x}$ and is denoted as $\Gamma_{G,i}(\mathbf{x})$. Note $D_G(\mathbf{x}) = \sum_{i=1}^{d} \Gamma_{G,i}(\mathbf{x})$.*
- *The* thresholded degree *of $\mathbf{x}$ in $G$ is the number of coordinates $i \in [d]$ with $\Gamma_{G,i}(\mathbf{x}) > 0$ and is denoted as $\Phi_G(\mathbf{x})$.*

*Whenever $G$ is clear from context, for brevity, we remove it from the subscript.*

Note that $\Phi(\mathbf{x})$ is an integer between $0$ and $d$, $\Gamma_i(\mathbf{x})$ is an integer between $0$ and $(n-1)$, and $D(\mathbf{x})$ is an integer between $0$ and $(n-1)d$. We next define the following parameters of a violation subgraph $G$.

**Definition 3.3.** *Consider a violation subgraph $G = (\boldsymbol{X}, \boldsymbol{Y}, E)$.*
- *$D(\boldsymbol{X})$ is the maximum degree of a vertex in $\boldsymbol{X}$, that is, $D(\boldsymbol{X}) = \max_{\mathbf{x} \in \boldsymbol{X}} D(\mathbf{x})$.*
- *For $i \in [d]$, $\Gamma_i(\boldsymbol{X})$ is the maximum $i$-degree in $\boldsymbol{X}$, that is, $\Gamma_i(\boldsymbol{X}) = \max_{\mathbf{x} \in \boldsymbol{X}} \Gamma_i(\mathbf{x})$.*
- *$\Gamma(\boldsymbol{X})$ is the maximum value of $\Gamma_i(\boldsymbol{X})$, that is, $\Gamma(\boldsymbol{X}) = \max_{i=1}^{d} \Gamma_i(\boldsymbol{X})$.*
- *$\Phi(\boldsymbol{X})$ is the maximum* thresholded *degree in $\boldsymbol{X}$, that is, $\Phi(\boldsymbol{X}) = \max_{\mathbf{x} \in \boldsymbol{X}} \Phi(\mathbf{x})$.*
- *$m(G)$ is the number of edges in $G$.*

*(We analogously define these parameters for $\boldsymbol{Y}$.)*

We recall the notion of *thresholded influence* of a function $f : [n]^d \to \{0, 1\}$ as defined in [BCS23, BKKM23]. For any $\mathbf{x} \in [n]^d$ and $i \in [d]$, $\Phi_f(\mathbf{x}; i)$ is the indicator for the existence of a violating $i$-edge incident to $\mathbf{x}$. The thresholded influence of $f$ at $\mathbf{x}$ is $\Phi_f(\mathbf{x}) = \sum_{i=1}^{d} \Phi_f(\mathbf{x}; i)$. We use the same Greek letter $\Phi$ both for thresholded influence and thresholded degree. In the graph $G_0 = (\boldsymbol{X}_0, \boldsymbol{Y}_0, E)$ consisting of *all* violating edges of the fully augmented hypergrid, $\Phi_f(\mathbf{x})$ is indeed $\Phi_{G_0}(\mathbf{x})$.

For applications to monotonicity testing, we require *colored/robust* versions of the thresholded influence. For hypercubes this was suggested by [KMS18], and for hypergrids this was generalized by [BCS23]. Let $\chi : E \to \{0,1\}$ be an *arbitrary* coloring of all the edges of the fully augmented hypergrid to 0 or 1. Given a point $\mathbf{x}$ and $i \in [d]$, $\Phi_{f,\chi}(\mathbf{x}; i)$ is the indicator of a violating $i$-edge $e$ incident to $\mathbf{x}$ with $\chi(e) = f(\mathbf{x})$. The colored thresholded influence of $\mathbf{x}$ wrt $\chi$ is simply $\Phi_{f,\chi}(\mathbf{x}) = \sum_{i=1}^{d} \Phi_{f,\chi}(\mathbf{x}; i)$. The Talagrand objective of $f$ is defined as

$$\mathrm{Tal}(f) := \min_{\chi: E \to \{0,1\}} \sum_{\mathbf{x} \in [n]^d} \sqrt{\Phi_{f,\chi}(\mathbf{x})}.$$

The main result of [BCS23] is the following.

**Theorem 3.4** (Theorem 1.4, [BCS23]). *If $f\colon [n]^d \to \{0,1\}$ is $\varepsilon$-far from monotone, then* $\mathrm{Tal}(f) = \Omega(\frac{\varepsilon n^d}{\log n})$.

We stress that the RHS above only loses a $\log n$ factor, which allows for domain reduction (setting $n = \mathrm{poly}(d)$). This is what yields the nearly optimal $\sqrt{d}$ dependence and independence on $n$ in the tester query complexity.

We extend the definition of $\mathrm{Tal}(f)$ to arbitrary violation subgraphs as follows. Given a violation subgraph $G = (\boldsymbol{X}, \boldsymbol{Y}, E)$ and a bicoloring $\chi : E \to \{0,1\}$ of its edges, for $\mathbf{z} \in \boldsymbol{X} \cup \boldsymbol{Y}$ and $i \in [d]$ let $\Phi_{G,\chi}(\mathbf{z}; i) = 1$ if there is a violating $i$-edge $e \in E(G)$ incident to $\mathbf{z}$ such that $\chi(e) = f(\mathbf{z})$, and $\Phi_{G,\chi}(\mathbf{z}; i) = 0$ otherwise. Define $\Phi_{G,\chi}(\mathbf{x}) = \sum_{i=1}^{d} \Phi_{G,\chi}(\mathbf{x}; i)$. Note, if $\chi \equiv 1$, that is every edge is colored 1, then $\Phi_{G,\chi}(\mathbf{x}) = \Phi_G(\mathbf{x})$ for $\mathbf{x} \in \boldsymbol{X}$ and $\Phi_{G,\chi}(\mathbf{y}) = 0$ for all $\mathbf{y} \in \boldsymbol{Y}$. Similarly, if $\chi \equiv 0$, then $\Phi_{G,\chi}(\mathbf{y}) = \Phi_G(\mathbf{y})$ for $\mathbf{y} \in \boldsymbol{Y}$ and $\Phi_{G,\chi}(\mathbf{x}) = 0$ for $\mathbf{x} \in \boldsymbol{X}$.

**Definition 3.5.** *Given a violation subgraph $G = (\boldsymbol{X}, \boldsymbol{Y}, E)$, we define* $\mathrm{Tal}(G) := \min_{\chi} \sum_{\mathbf{z} \in \boldsymbol{X} \cup \boldsymbol{Y}} [\sqrt{\Phi_{G,\chi}(\mathbf{z})}]$, *where the min is taken over all edge bicolorings $\chi : E(G) \to \{0,1\}$.*

If $G_0$ is the subgraph of all violations in the fully augmented hypergrid, then Theorem 3.4 states $\mathrm{Tal}(G_0) = \Omega(\varepsilon n^d / \log n)$. We make a couple of observations.

**Observation 3.6.** *For any violation subgraph $G = (\boldsymbol{X}, \boldsymbol{Y}, E)$,*

- $D(\boldsymbol{X}) \le \Gamma(\boldsymbol{X})\Phi(\boldsymbol{X})$ *and* $D(\boldsymbol{Y}) \le \Gamma(\boldsymbol{Y})\Phi(\boldsymbol{Y})$.
- $m(G) \ge \mathrm{Tal}(G)$.

*Proof.* For any $\mathbf{x} \in \boldsymbol{X}$, we have $D(\mathbf{x}) = \sum_{i=1}^{d} \Gamma_i(\mathbf{x}) = \sum_{i:\Gamma_i(\mathbf{x})>0} \Gamma_i(\mathbf{x}) \le (\max_i \Gamma_i(\mathbf{x})) \cdot \Phi(\mathbf{x}) \le \Gamma(\boldsymbol{X})\Phi(\boldsymbol{X})$. The proof is analogous for $\boldsymbol{Y}$. For the second bullet, observe that $m(G) = \sum_{\mathbf{x} \in \boldsymbol{X}} D(\mathbf{x}) \ge \sum_{\mathbf{x} \in \boldsymbol{X}} \Phi(\mathbf{x}) \ge \sum_{\mathbf{x} \in \boldsymbol{X}} \sqrt{\Phi(\mathbf{x})} = \sum_{\mathbf{z} \in \boldsymbol{X} \cup \boldsymbol{Y}} \sqrt{\Phi_{G,\chi \equiv 1}(\mathbf{z})} \ge \mathrm{Tal}(G)$. $\square$

**Remark 3.7.** *Throughout the remainder of the paper, we consider $d$ to be at least a large constant and fix $\delta > \frac{1}{\log\log nd}$. As a result, we use bounds such as "$d^{\delta} \ge \mathrm{poly}\log d$" or "$d - C\sqrt{d} \ge d/3$" without explicitly reminding the reader that $d$ is large. We use $O(\delta)$ to denote $C \cdot \delta$ for some unspecified, but fixed constant $C$.*

## 3.2 Equivalent Formulations of the Random Walk Distribution

Recall the random walk distribution described in Definition 2.1. It is useful to think of this walk as first sampling a random hypercube and then taking a random walk on that hypercube. The following definition describes the appropriate distribution over sub-hypercubes in $[n]^d$.

**Definition 3.8** (Hypercube Distribution)**.** *We define the following distribution $\mathbb{H}_{n,d}$ over sub-hypercubes in $[n]^d$. For each coordinate $i \in [d]$:*

1. *Choose $q_i \in_R \{1, 2, \ldots, \log n\}$ uniformly at random.*
2. *Choose a uniform random interval $I_i$ of size $2^{q_i}$ in $\mathbb{Z}_n$.*
3. *Choose a uniform random pair $a_i < b_i$ from $I_i$.*

*Output $\boldsymbol{H} = \prod_{i=1}^{d}\{a_i, b_i\}$. When $n$ and $d$ are clear from context, we abbreviate $\mathbb{H} = \mathbb{H}_{n,d}$.*

It will also be useful for us to think of our random walk distribution as first sampling $\mathbf{x} \in_R [n]^d$, then sampling a random hypercube which contains $\mathbf{x}$, and then taking a random walk from $\mathbf{x}$ in that hypercube. The appropriate distribution over hypercubes containing a point $\mathbf{x}$ is defined as follows.

**Definition 3.9** (Conditioned Hypercube Distribution)**.** *Given $\mathbf{x} \in [n]^d$, we define the conditioned sub-hypercube distribution $\mathbb{H}_{n,d}(\mathbf{x})$ as follows. For each $i \in [d]$:*

1. *Choose $q_i \in_R \{1, 2, \ldots, \log n\}$ uniformly at random.*
2. *Choose a uniform random interval $I_i$ in $\mathbb{Z}_n$ of size $2^{q_i}$ such that $\mathbf{x}_i \in I_i$.*
3. *Choose a uniform random $c_i \in_R I_i \setminus \{\mathbf{x}_i\}$.*
4. *Set $a_i = \min(\mathbf{x}_i, c_i)$ and $b_i = \max(\mathbf{x}_i, c_i)$.*

*Output $\boldsymbol{H} = \prod_{i=1}^{d}\{a_i, b_i\}$. When $n$ and $d$ are clear from context we will abbreviate $\mathbb{H}(\mathbf{x}) = \mathbb{H}_{n,d}(\mathbf{x})$.*

The random walk distribution in a hypercube $\boldsymbol{H}$ is defined as follows.

**Definition 3.10** (Hypercube Walk Distribution)**.** *For a hypercube $\boldsymbol{H} = \prod_{i=1}^{d}\{a_i, b_i\}$, a point $\mathbf{x} \in \boldsymbol{H}$, and a walk length $\tau$, we define the upward random walk distribution $\mathcal{U}_{\boldsymbol{H},\tau}(\mathbf{x})$ over points $\mathbf{y} \in \boldsymbol{H}$ as follows.*

1. *Pick a uniform random subset $R \subseteq [d]$ of $\tau$ coordinates.*
2. *Generate $\mathbf{y}$ as follows. For every $r \in [d]$, if $r \in R$ and $\mathbf{x}_r = a_r$, set $\mathbf{y}_r = b_r$. Else, set $\mathbf{y}_r = \mathbf{x}_r$.*

*Analogously, the downward random walk distribution $\mathcal{D}_{\boldsymbol{H},\tau}(\mathbf{x})$ is defined precisely as above, but instead in step 2 if $r \in R$ and $\mathbf{x}_r = b_r$, we set $\mathbf{y}_r = a_r$, and otherwise $\mathbf{y}_r = \mathbf{x}_r$.*

We observe that the following walk distributions are equivalent and defer the proof to the appendix §A.1.

**Fact 3.11.** *The following three distributions over pairs $(\mathbf{x}, \mathbf{y}) \in [n]^d \times [n]^d$ are all equivalent.*

1. *$\mathbf{x} \in_R [n]^d$, $\mathbf{y} \sim \mathcal{U}_\tau(\mathbf{x})$.*
2. *$\boldsymbol{H} \sim \mathbb{H}$, $\mathbf{x} \in_R \boldsymbol{H}$, $\mathbf{y} \sim \mathcal{U}_{\boldsymbol{H},\tau}(\mathbf{x})$.*
3. *$\mathbf{x} \in_R [n]^d$, $\boldsymbol{H} \sim \mathbb{H}(\mathbf{x})$, $\mathbf{y} \sim \mathcal{U}_{\boldsymbol{H},\tau}(\mathbf{x})$.*

*The analogous three distributions defined using downward random walks are also equivalent.*

It is also convenient to define the shift distribution for hypercubes.

**Definition 3.12** (Shift Distributions for Hypercube Walks)**.** *Given a hypercube $\boldsymbol{H}$, the up-shift distribution from $\mathbf{x} \in \boldsymbol{H}$, denoted $\mathcal{US}_{\boldsymbol{H},\tau}(\mathbf{x})$ is the distribution of $\mathbf{x}' - \mathbf{x}$, where $\mathbf{x}' \sim \mathcal{U}_{\boldsymbol{H},\tau}(\mathbf{x})$. The down-shift distribution from $\mathbf{y} \in \boldsymbol{H}$, denoted $\mathcal{DS}_{\boldsymbol{H},\tau}(\mathbf{y})$ is the distribution of $\mathbf{y} - \mathbf{y}'$, where $\mathbf{y}' \sim \mathcal{DS}_{\boldsymbol{H},\tau}(\mathbf{y})$.*

### 3.3 Influence and Persistence

We define the following notion of influence for our random walk distribution Definition 2.1.

**Definition 3.13.** *The total and negative influences of* $f \colon [n]^d \to \{0,1\}$ *are defined as follows.*

- $\widetilde{I}_f = \mathbb{E}_{\mathbf{x} \in [n]^d} \left[ d \cdot \Pr_{\mathbf{y} \sim \mathcal{U}_1(\mathbf{x})}[f(\mathbf{x}) \neq f(\mathbf{y})] \right]$
- $\widetilde{I}_f^- = \mathbb{E}_{\mathbf{x} \in [n]^d} \left[ d \cdot \Pr_{\mathbf{y} \sim \mathcal{U}_1(\mathbf{x})}[f(\mathbf{x}) > f(\mathbf{y})] \right]$

The probability of the tester (Alg. 1) finding a violation in step (2b) when $\tau = 1$ is precisely $\widetilde{I}_f^-/d$. Recall the definition of the distribution $\mathbb{H}$ in Definition 3.8. For brevity, for a hypercube $\boldsymbol{H} = \prod_{i=1}^d \{a_i, b_i\}$ sampled from $\mathbb{H}$, we abbreviate $I_{\boldsymbol{H}} := I_{f|_{\boldsymbol{H}}}$ and $I_{\boldsymbol{H}}^- := I_{f|_{\boldsymbol{H}}}^-$. That is, if $f(\mathbf{x}) = 1$, then $I_{\boldsymbol{H}}(\mathbf{x})$ is the number of coordinates $i$ for which $\mathbf{x}_i = a_i$, and $f(\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}, b_i, \mathbf{x}_{i+1}, \ldots, \mathbf{x}_d) = 0$, and if $f(\mathbf{x}) = 0$, then $I_{\boldsymbol{H}}(\mathbf{x}) = 0$. Then, $I_{\boldsymbol{H}} = \mathbb{E}_{\mathbf{x} \in \boldsymbol{H}}[I_{\boldsymbol{H}}(\mathbf{x})]$. The definition is analogous for $I_{\boldsymbol{H}}^-$.

**Claim 3.14.** $\widetilde{I}_f = \mathbb{E}_{\boldsymbol{H} \sim \mathbb{H}}[I_{\boldsymbol{H}}]$ *and* $\widetilde{I}_f^- = \mathbb{E}_{\boldsymbol{H} \sim \mathbb{H}}[I_{\boldsymbol{H}}^-]$.

*Proof.* By Fact 3.11, the distribution $(\mathbf{x} \in_R [n]^d, \mathbf{y} \sim \mathcal{U}_1(\mathbf{x}))$ is equivalent to first sampling $\boldsymbol{H} \sim \mathbb{H}$, then sampling $(\mathbf{x} \in_R \boldsymbol{H}, \mathbf{y} \sim \mathcal{U}_{\boldsymbol{H},1}(\mathbf{x}))$. Recalling Definition 3.10, observe that $\Pr_{\mathbf{y} \sim \mathcal{U}_{\boldsymbol{H},1}(\mathbf{x})}[f(\mathbf{x}) \neq f(\mathbf{y})] = I_{\boldsymbol{H}}(\mathbf{x})/d$. Putting these observations together yields

$$\widetilde{I}_f = \mathbb{E}_{\mathbf{x} \in [n]^d} \left[ d \cdot \Pr_{\mathbf{y} \sim \mathcal{U}_1(\mathbf{x})}[f(\mathbf{x}) \neq f(\mathbf{y})] \right] = \mathbb{E}_{\boldsymbol{H} \sim \mathbb{H}} \mathbb{E}_{\mathbf{x} \in \boldsymbol{H}} \left[ I_{\boldsymbol{H}}(\mathbf{x}) \right] = \mathbb{E}_{\boldsymbol{H} \sim \mathbb{H}}[I_{\boldsymbol{H}}]$$

An analogous argument proves the statement for negative influence. □

The following claim states that if the normal influence is (very) large, then so is the negative influence. This is a simple generalization of, and indeed easily follows from, Theorem 9.1 in [KMS18]. The proof can be found in §A.2.

**Claim 3.15.** *If* $\widetilde{I}_f > 9\sqrt{d}$*, then* $\widetilde{I}_f^- > \sqrt{d}$*.*

Next, we define the notion of persistent points. This is similar to that in [KMS18] with a parameterization that we need for our purpose.

**Definition 3.16.** *Given a point* $\mathbf{x} \in [n]^d$*, a walk length* $\tau$*, and a parameter* $\beta \in (0,1)$*, we say that* $\mathbf{x}$ *is* $(\tau, \beta)$*-up-persistent if*

$$\Pr_{\mathbf{y} \sim \mathcal{U}_\tau(\mathbf{x})}[f(\mathbf{y}) \neq f(\mathbf{x})] \leq \beta.$$

*Similarly,* $\mathbf{x}$ *is called* $(\tau, \beta)$*-down-persistent if the above bound holds when* $\mathbf{y}$ *is drawn from the downward walk distribution,* $\mathcal{D}_\tau(\mathbf{x})$*. If both bounds hold, then we call* $\mathbf{x}$ $(\tau, \beta)$*-persistent.*

The following claim upper bounds the fraction of non-persistent points. This is a generalization of Lemma 9.3 in [KMS18]. The proof is deferred to §A.2.

**Claim 3.17.** *If* $\widetilde{I}_f \leq 9\sqrt{d}$*, then the fraction of vertices that are not* $(\tau, \beta)$*-persistent is at most* $C_{per} \frac{\tau}{\beta\sqrt{d}}$ *where* $C_{per}$ *is a universal constant.*

### 3.4 The Middle Layers, Typical Points, and Walk Reversibility

All proofs in this section are deferred to §A.3.

**Definition 3.18.** *In a hypercube* $\{0,1\}^d$*, the* $c$-middle layers *consist of all points with Hamming weight in the range* $[d/2 \pm \sqrt{4cd\log(d/\varepsilon)}]$*. Given a* $d$*-dimensional hypercube* $\boldsymbol{H}$*, we let* $\boldsymbol{H}_c \subseteq \boldsymbol{H}$ *denote the* $c$*-middle layers of* $\boldsymbol{H}$.

We state a bound on the number of points in the hypercube which lie in the middle layers. This follows from a standard Chernoff bound argument.

**Claim 3.19.** *For a* $d$*-dimensional hypercube* $\boldsymbol{H}$ *and* $c \geq 1$*, we have* $|\boldsymbol{H}_c| \geq (1-(\varepsilon/d)^c)\cdot 2^d$.

We now define the notion of typical points in $[n]^d$. Recall the distribution $\mathbb{H}_{n,d}$ (Definition 3.8) over random sub-hypercubes in $[n]^d$ and the distribution $\mathbb{H}_{n,d}(\mathbf{x})$ (Definition 3.8) over random sub-hypercubes in $[n]^d$ that contain $\mathbf{x}$. A point $\mathbf{x}$ is $c$*-typical* if for most sub-hypercubes containing $\mathbf{x}$, the point $\mathbf{x}$ is present in their $c$-middle layers.

**Definition 3.20** (Typical Points)**.** *Given* $c \geq 1$*, a point* $\mathbf{x} \in [n]^d$ *is called* $c$-typical *if*

$$\Pr_{\boldsymbol{H}\sim\mathbb{H}(\mathbf{x})}[\mathbf{x} \in \boldsymbol{H}_c] \geq 1-(\varepsilon/d)^5.$$

**Claim 3.21** (Most Points are Typical)**.** *For any* $\varepsilon \in (0,1)$ *and* $c \geq 6$*,*

$$\Pr_{\mathbf{x}\in_R [n]^d}[\mathbf{x} \text{ is } c\text{-typical}] \geq 1-(\varepsilon/d)^{c-5}.$$

Intuitively, a short random walk from a typical point will always lead to point that is almost as typical. This is formalized as follows.

**Claim 3.22** (Translations of Typical Points)**.** *Suppose* $\mathbf{x} \in [n]^d$ *is* $c$*-typical. Then for a walk length* $\tau \leq \sqrt{d}$*, every point* $\mathbf{x}' \in \mathit{supp}(\mathcal{U}_\tau(\mathbf{x})) \cup \mathit{supp}(\mathcal{D}_\tau(\mathbf{x}))$ *is* $(c+\frac{\tau}{\sqrt{d}})$*-typical.*

Recall the three equivalent ways of expressing the walk distribution in Fact 3.11. We define the random walk probabilities only on points in the middle layers. This setup allows for the approximate reversibility of Lemma 3.24.

**Definition 3.23.** *Consider two vertices* $\mathbf{x} \prec \mathbf{x}' \in [n]^d$ *and a walk length* $\tau$*. We define*

$$p_{\mathbf{x},\tau}(\mathbf{x}') = \mathbb{E}_{\boldsymbol{H}\sim\mathbb{H}(\mathbf{x})}\left[\mathbf{1}\left(\mathbf{x} \in \boldsymbol{H}_{100} \;\wedge\; \mathbf{x}' \in \boldsymbol{H}_{100}\right)\cdot \Pr_{\mathbf{z}\sim\mathcal{U}_{\boldsymbol{H},\tau}(\mathbf{x})}[\mathbf{z}=\mathbf{x}']\right] \tag{1}$$

*to be the probability of reaching* $\mathbf{x}'$ *by a random walk from* $\mathbf{x}$*, only counting the contribution when the random walk is taken on a hypercube that contains* $\mathbf{x}$ *and* $\mathbf{x}'$ *in the* 100*-middle layers. We analogously define* $p_{\mathbf{x}',\tau}(\mathbf{x})$ *using the downward random walk distribution in* $\boldsymbol{H}$.

Consider $\mathbf{x} \prec \mathbf{x}'$ are two points in the middle layers. The following lemma asserts that the probability of reaching from $\mathbf{x}$ to $\mathbf{x}'$ via an upward walk of length $\ll \sqrt{d}$ is similar to the probability of reaching from $\mathbf{x}'$ to $\mathbf{x}$ via downward walk of the same length.

**Lemma 3.24** (Reversibility Lemma)**.** *For any points* $\mathbf{x} \prec \mathbf{x}' \in [n]^d$ *and walk length* $\ell \leq \sqrt{d}/\log^5(d/\varepsilon)$*, we have*

$$p_{\mathbf{x},\ell}(\mathbf{x}') = (1 \pm \log^{-3} d)\, p_{\mathbf{x}',\ell}(\mathbf{x}).$$

# 4 Red Edges, Blue Edges, and Nice Subgraphs

We now set the stage to prove Theorem 2.4. The first definition is that of *mostly-zero-below* points. These are points from which a downward random walk (Definition 2.1) leads to a point where the function evaluates to 0 with high probability.

**Definition 4.1.** *A point* $\mathbf{z}$ *is called* $\ell$-mostly-zero-below, *or* $\ell$-mzb, *if* $\Pr_{\mathbf{z}'\sim\mathcal{D}_\ell(\mathbf{z})}[f(\mathbf{z}')=0]\geq 0.9$.

To appreciate the utility of $\ell$-mzb points, consider the following scenario. Suppose $\mathbf{x}$ is a point with $f(\mathbf{x})=1$ and is $(\ell,\beta)$-down-persistent (Definition 3.16) for some small $\beta$. Next suppose an upward random walk from $\mathbf{x}$ reaches an $\ell$-mzb point $\mathbf{z}$. Then, we claim that Step 4 of Alg. 1 would succeed with constant probability in finding a violated edge. An $\ell$-length downward walk from $\mathbf{x}$, due to down-persistence, would lead to a $\mathbf{x}'$ with $f(\mathbf{x}')=1$ with probability at least $1-\beta$. The same $\ell$-length downward walk from $\mathbf{z}$ would lead to a $\mathbf{z}'$ with $f(\mathbf{z}')=0$ with $\geq 0.9$ probability, since $\mathbf{z}$ is mostly-zero-below. Since $(\mathbf{x},\mathbf{z})$ are comparable, so would be $(\mathbf{x}',\mathbf{z}')$. By a union bound, $(\mathbf{x}',\mathbf{z}')$ is a violation with probability at least $0.9-\beta$.

The next definition describes edges $(\mathbf{x},\mathbf{y})$ of the violation subgraph most of whose internal vertices lead to mzb-points via an upward random walk. Uncreatively, we call such edges *red*. Recall that $I(\mathbf{x},\mathbf{y})=\{\mathbf{z}\colon \mathbf{x}\preceq\mathbf{z}\preceq\mathbf{y}\}$ denotes the closed interval of points from $\mathbf{x}$ to $\mathbf{y}$.

**Definition 4.2.** *A violated edge* $(\mathbf{x},\mathbf{y})$ *is called/colored* red *for walk length* $\ell$ *if*

$$\Pr_{\mathbf{z}\in I(\mathbf{x},\mathbf{y})}\ \Pr_{\mathbf{z}'\sim\mathcal{U}_\ell(\mathbf{z})}[\mathbf{z}' \textit{ is } \ell\text{-}\mathsf{mzb}]\geq 0.01.$$

*When* $\ell$ *is clear by context, we call the edge red.*

There may be no $\ell$-mzb points for the lengths we choose, that is, a downward walk from any point leads to a point where the function evaluates to 1. In that case, Step 3 of Alg. 1 is poised to succeed; for any violating edge $(\mathbf{x},\mathbf{y})$, if we start from $\mathbf{y}$ then the downward walk should give a violation. This motivates the next definition which recognizes violated edges $(\mathbf{x},\mathbf{y})$ most of whose internal vertices lead to points where the function evaluates to 1 via a downward random walk. We call such edges *blue*.

**Definition 4.3.** *A violated edge* $(\mathbf{x},\mathbf{y})$ *is called/colored* blue *for walk length* $\ell$ *if*

$$\Pr_{\mathbf{z}\in I(\mathbf{x},\mathbf{y})}\ \Pr_{\mathbf{z}'\sim\mathcal{D}_\ell(\mathbf{z})}[f(\mathbf{z}')=1]\geq 0.01.$$

*When* $\ell$ *is clear by context, we simply call the edge blue.*

We note that a violating edge $(\mathbf{x},\mathbf{y})$ may be *both* red and blue, or perhaps more problematically, *neither* red nor blue. One of the key lemmas we prove is that we can get our hands on a violation subgraph with sufficiently many colored edges. If we have our hands on a large violation subgraph $G$ with few red edges (but has some other properties), then we can find another comparable sized violation subgraph $H$ all of whose edges are blue, and whose maximum degrees are bounded by those in $G$. The precise statement is given below. We defer the proof of this lemma to §7.

**Lemma 4.4** (Red/Blue Lemma). *Let* $G(\boldsymbol{X},\boldsymbol{Y},E)$ *be a violation subgraph and* $1\leq\ell\leq\sqrt{d}/\log^5(d/\varepsilon)$ *be a walk length such that the following hold.*

1. *At most half the edges are red for walk length* $\ell$.
2. *All vertices in* $\boldsymbol{X}\cup\boldsymbol{Y}$ *are* $(\ell,\log^{-5}d)$*-up-persistent.*

3. *All vertices in $\boldsymbol{X} \cup \boldsymbol{Y}$ are* 99*-typical.*

*Then there exists another violation subgraph $H(\boldsymbol{L}, \boldsymbol{R}, E')$ such that*

1. *All edges are blue for walk length $\ell$ and $m(H) \geq m(G)/7$.*
2. $\Gamma(\boldsymbol{L}) \leq \Gamma(\boldsymbol{X})$ *and* $\Gamma(\boldsymbol{R}) \leq \Gamma(\boldsymbol{Y})$.
3. $D(\boldsymbol{L}) \leq D(\boldsymbol{X})$ *and* $D(\boldsymbol{R}) \leq D(\boldsymbol{Y})$.

The next two definitions capture certain "nice" violation subgraphs consisting of either red or blue edges. In §5, we show that if either of these subgraphs exist then we can prove the tester works with the desired probability. In §6 we show that one of these subgraphs must exist. Recall, $\Phi_H(\mathbf{x})$ is the *thresholded degree* of $\mathbf{x}$ in the subgraph $H$ and $\delta > (\log \log nd)^{-1}$ is fixed (Remark 3.7).

**Definition 4.5** (($\sigma, \tau$)-nice red violation subgraph)**.** *Given a parameter $\sigma \in (0,1)$ and a walk length $\tau$, a violation subgraph $H(\boldsymbol{A}, \boldsymbol{B}, E)$ is called a $(\sigma, \tau)$-nice red violation subgraph if the following hold.*

*(a) All edges in $H$ are red for walk length $\tau - 1$.*

*(b) All vertices in $\boldsymbol{A}$ are $(\tau - 1, 0.6)$-down-persistent.*

*(c)* $\sigma \Phi_H(\mathbf{x}) \leq d^{1/2}$ *for all* $\mathbf{x} \in \boldsymbol{A}$.

*(d)* $\sigma \sum_{\mathbf{x} \in \boldsymbol{A}} \Phi_H(\mathbf{x}) \geq \varepsilon^2 \cdot n^d \cdot d^{-O(\delta)}$.

*(e)* $d^{1/2-O(\delta)} \geq \tau \geq \sigma \cdot d^{1/2-O(\delta)}$.

The first two conditions dictate that the subgraph is nice with respect to the length of the walk. In particular, the edges are red with respect to this length and furthermore the 1-vertices are down-persistent. As explained before the definition of red edges, this property is crucial for the success of Step 4 of Alg. 1. The fourth condition says that the total thresholded degree of the 1-vertices in $H$ is large. I.e. for an average vertex $\mathbf{x} \in \boldsymbol{A}$, there will be many coordinates $i$ for which there is an $i$-edge in $H$ incident to $\mathbf{x}$. The third condition says that the max thresholded degree of vertices in $\boldsymbol{A}$ is not too large and so the total thresholded degree from the fourth condition must be somewhat spread amongst the vertices in $\boldsymbol{A}$. The final condition shows that the length of the walk is large compared to $\sigma$. Note, if $\sigma = \Theta(1)$ and the third bullet point's right hand side was 1 instead of $\sqrt{d}$, we would be in the case of a large matching of violated edges, which was the "simple case" discussed in §1.2.

The next definition is the analogous case of blue edges. When this type of subgraph exists we argue that Step 3 of Alg. 1 succeeds. Note that Step 3 is the downward path test (without a shift) and so we don't need a persistence property like condition (b) in the previous definition. This definition has the same conditions on the thresholded degree as the previous definition, but with respect to the 0-vertices of the subgraph.

**Definition 4.6** (($\sigma, \tau$)-nice blue violation subgraph)**.** *Given a parameter $\sigma \in (0,1)$ and a walk length $\tau$, a violation subgraph $H(\boldsymbol{A}, \boldsymbol{B}, E)$ is called a $(\sigma, \tau)$-nice blue violation subgraph if the following hold.*

*(a) All edges in $H$ are blue for walk length $\tau - 1$.*

*(b)* $\sigma \Phi_H(\mathbf{y}) \leq d^{1/2}$ *for all* $\mathbf{y} \in \boldsymbol{B}$.

*(c)* $\sigma \sum_{\mathbf{y} \in \boldsymbol{B}} \Phi_H(\mathbf{y}) \geq \varepsilon^2 \cdot n^d \cdot d^{-O(\delta)}$.

*(d)* $d^{1/2-O(\delta)} \geq \tau \geq \sigma \cdot d^{1/2-O(\delta)}$.

The following lemma captures the utility of the above definitions. It's proof can be found in §5.

**Lemma 4.7** (Nice Subgraphs and Random Walks)**.** *Suppose for a power of two $\tau \geq 2$, there exists a $(\sigma, \tau)$-nice red subgraph or a $(\sigma, \tau)$-nice blue subgraph. Then Alg. 1 finds a violating pair, and thus rejects $f$, with probability at least $\varepsilon^2 \cdot d^{-(1/2+O(\delta))}$.*

The following lemma shows that one of the two nice subgraphs always exists. It's proof can be found in §6.

**Lemma 4.8** (Existence of nice subgraphs)**.** *Let $n, \varepsilon^{-1} \leq \mathrm{poly}(d)$. Suppose $f\colon [n]^d \to \{0,1\}$ is $\varepsilon$-far from monotone and $\widetilde{I}_f \leq 9\sqrt{d}$. Let $\delta > \frac{1}{\log\log nd}$ be a parameter. There exists $0 < \sigma_1 \leq \sigma_2 < 1$, a violation subgraph $H(\boldsymbol{A}, \boldsymbol{B}, E)$, and a power of two $\tau \geq 2$, such that either $H$ is a $(\sigma_1, \tau)$-nice red subgraph or a $(\sigma_2, \tau)$-nice blue subgraph.*

## 5 Tester Analysis

In this section we prove Theorem 2.4. First, in §5.1 we prove Lemma 4.7 which is the main tester analysis. Then in §5.2 we combine Lemma 4.7, Lemma 4.8 (which will be proven in §6), and Claim 3.15 to prove Theorem 2.4.

### 5.1 Main Analysis: Proof of Lemma 4.7

There are two cases depending on whether we have a nice red subgraph or a nice blue subgraph. In Case 1, Step 4 of Alg. 1 proves the lemma while in Case 2, Step 3 of Alg. 1 proves the lemma. The proofs are similar, but we provide both for completeness.

#### 5.1.1 Case 1: $H$ is a $(\sigma, \tau)$-nice red subgraph

Since $\tau$ is a power of 2, the tester in Alg. 1 chooses it with probability $\log^{-1} d$. Thus, in the rest of the analysis we will condition on this event.

Given $\mathbf{x} \in \boldsymbol{A}$, let $C_{\mathbf{x}} \subseteq [d]$ denote the set of coordinates for which $\mathbf{x}$ has an outgoing edge in $H$. Note $|C_{\mathbf{x}}| = \Phi_H(\mathbf{x})$. Recall the upward path + downward shift test described in Step 4 of Alg. 1 and the walk distribution $\mathcal{U}_{\tau-1}(\mathbf{x})$ defined in Definition 2.1. We first lower bound the probability that $\mathbf{x} \in \boldsymbol{A}$ and $R \cap C_{\mathbf{x}} \neq \emptyset$ where $\mathbf{x}$ is chosen uniformly by the tester and $R \subseteq [d]$ is a random set of $\tau$ coordinates. Let $\mathcal{E}_1$ denote this event. The main calculation is to lower bound the probability of this event as follows.

$$\Pr[\mathcal{E}_1] = \frac{1}{n^d} \sum_{\mathbf{x} \in \boldsymbol{A}} \Pr[R \cap C_{\mathbf{x}} \neq \emptyset] \geq \frac{1}{n^d} \sum_{\mathbf{x} \in \boldsymbol{A}} \left[1 - \left(1 - \frac{|C_{\mathbf{x}}|}{d}\right)^{\tau}\right] \geq \frac{1}{n^d} \sum_{\mathbf{x} \in \boldsymbol{A}} \left[1 - \exp\left(-\frac{\tau |C_{\mathbf{x}}|}{d}\right)\right]$$

The RHS can only decrease if we replace $\tau$ with its lower bound (Definition 4.5, (e)) of $\sigma \cdot d^{1/2-O(\delta)}$. Also, observe that $\frac{\sigma d^{1/2-O(\delta)} |C_{\mathbf{x}}|}{d} = \frac{\sigma \Phi_H(\mathbf{x})}{d^{1/2+O(\delta)}} \leq 1$ using our upper bound, $\sigma \Phi_H(\mathbf{x}) \leq d^{1/2}$ (Definition 4.5, (c)). Now, using $e^{-x} \leq 1 - \frac{x}{2}$ for $x \leq 1$, the exponential term in the RHS is at most $1 - \frac{\sigma \Phi_H(\mathbf{x})}{2d^{1/2+O(\delta)}}$, yielding

$$\Pr[\mathcal{E}_1] \geq \frac{\sigma}{2d^{1/2+O(\delta)}} \cdot \frac{1}{n^d} \sum_{\mathbf{x} \in \boldsymbol{A}} \Phi_H(\mathbf{x}) \underbrace{\geq}_{\text{(Definition 4.5, (d))}} \frac{\varepsilon^2}{d^{1/2+O(\delta)}} \tag{2}$$

The event $\mathcal{E}_1$ asserts that the tester has chosen a point $\mathbf{x} \in \boldsymbol{A}$ and there is at least one $r \in R$ for which there exists a red edge $(\mathbf{x}, \mathbf{x} + a\mathbf{e}_r) \in E$ for some integer $a > 0$ in the subgraph $H$. Fix the smallest such $r \in R \cap C_{\mathbf{x}}$ and the corresponding edge in $H$.

Recall the random walk process in Definition 2.1. We define the following good events.

- $\mathcal{E}_2$: Step (2a) chooses $q_r$ satisfying: if $a \le n/4$, then $2^{q_r} \in [2a, 4a]$; if $a > n/4$, then $2^{q_r} = n$.
- $\mathcal{E}_3$: Step (2b) chooses the interval $I_r \supseteq [\mathbf{x}_r, \mathbf{x}_r + a]$.
- $\mathcal{E}_4$: Step (2c) chooses $c_r$ uniformly[5] from $[\mathbf{x}_r, \mathbf{x}_r + a]$.
- $\mathcal{E}_5$: $\mathbf{y}$ is $(\tau - 1)$-mostly-zero-below as per Definition 4.1.
- $\mathcal{E}_6$: $f(\mathbf{y} - \mathbf{s}) = 0$ for $\mathbf{s}$ chosen in Step 4 of Alg. 1 from $\mathcal{DS}_{\tau-1}(\mathbf{x})$.
- $\mathcal{E}_7$: $f(\mathbf{x} - \mathbf{s}) = 1$ for $\mathbf{s}$ chosen in Step 4 of Alg. 1 from $\mathcal{DS}_{\tau-1}(\mathbf{x})$.

Firstly, note that $\Pr[\mathcal{E}_2] = \log^{-1} n$ for both cases of the edge length, $a$. Now, suppose $a \le n/4$. Then, $\Pr[\mathcal{E}_3 \mid \mathcal{E}_2] \ge 1/2$ by the condition $q_r \ge 2a$ and $\Pr[\mathcal{E}_4 \mid \mathcal{E}_2, \mathcal{E}_3] \ge 1/4$ by the condition $q_r \le 4a$. If $a > n/4$, then $\Pr[\mathcal{E}_3 \mid \mathcal{E}_2] = 1$, since in this case $I_r = [n]$ and again $\Pr[\mathcal{E}_4 \mid \mathcal{E}_2, \mathcal{E}_3] \ge 1/4$ since $[\mathbf{x}_r, \mathbf{x}_r + a]$ is at least a fourth of the entire line, $[n]$.

Now, since the edge $(\mathbf{x}, \mathbf{x}+a\mathbf{e}_r)$ is red (Definition 4.2) for walk length $\tau-1$, we have $\Pr[\mathcal{E}_5 \mid \mathcal{E}_4] \ge 0.01$.

Since $\mathbf{y}$ is $(\tau - 1)$-mostly-zero-below, if we sample $\mathbf{s}'$ from $\mathcal{DS}_{\tau-1}(\mathbf{y})$ we get $f(\mathbf{y} - \mathbf{s}') = 0$ with probability $\ge 0.9$. Now note that $\mathcal{DS}_{\tau-1}(\mathbf{y})$ and $\mathcal{DS}_{\tau-1}(\mathbf{x})$ differ only when the set $R \subseteq [d]$ chosen in Definition 2.1 contains a coordinate in $\mathrm{supp}(\mathbf{y} - \mathbf{x})$. Since $|\mathrm{supp}(\mathbf{y} - \mathbf{x})| \le \tau$, $|R| \le \tau$, and $\tau = o(\sqrt{d})$, we have $\Pr_R[R \cap \mathrm{supp}(\mathbf{y} - \mathbf{x}) \neq \emptyset] \le \tau^2/d = o(1)$. Therefore, when $\mathbf{s}$ is drawn from $\mathcal{DS}_{\tau-1}(\mathbf{x})$, we get $f(\mathbf{y} - \mathbf{s}) = 0$ with probability $\ge 0.9(1 - o(1)) \ge 0.8$. That is, $\Pr[\mathcal{E}_6 \mid \mathcal{E}_5] \ge 0.8$.

Finally, all points in $\boldsymbol{A}$ are $(\tau - 1, 0.6)$-down-persistent (Definition 3.16) and so $\Pr[\mathcal{E}_7 \mid \mathbf{x} \in A] \ge 0.4$. Now, let's put everything together. The final success probability of the tester is at least $\Pr[\mathcal{E}_6 \wedge \mathcal{E}_7]$, which by a union bound and the reasoning above, is at least

$$\begin{aligned} &(1 - \Pr[\neg\mathcal{E}_6 \mid \mathcal{E}_5] - \Pr[\neg\mathcal{E}_7 \mid \mathbf{x} \in A]) \cdot \Pr\left[\bigwedge_{i=1}^{5} \mathcal{E}_i\right] \\ &\ge (1 - 0.2 - 0.6) \cdot \frac{\varepsilon^2}{d^{1/2+O(\delta)}} \cdot \frac{1}{\log n} \cdot \frac{1}{2} \cdot \frac{1}{4} \cdot \frac{1}{100} \ge \frac{\varepsilon^2}{d^{1/2+O(\delta)}} \end{aligned}$$

where in the last inequality we used $n \le \mathrm{poly}(d)$. This completes the proof when the nice subgraph is red.

#### 5.1.2 Case 2: $H$ is a $(\sigma, \tau)$-nice blue subgraph

As in Case 1, since $\tau$ is a power of 2, the tester in Alg. 1 chooses it with probability $\log^{-1} d$. Thus, in the rest of the analysis we will condition on this event. Given $\mathbf{y} \in \boldsymbol{B}$, let $C_{\mathbf{y}} \subseteq [d]$ denote the set of coordinates for which $\mathbf{y}$ has an incoming edge in $H$. Note $|C_{\mathbf{y}}| = \Phi_H(\mathbf{y})$. Recall the downward path tester described in Step 3 of Alg. 1 and the walk distribution $\mathcal{D}_{\tau-1}(\mathbf{y})$ defined in Definition 2.1. We first lower bound the probability that $\mathbf{y} \in \boldsymbol{B}$ and $R \cap C_{\mathbf{y}} \neq \emptyset$ where $\mathbf{y}$ is chosen uniformly by the tester and $R \subseteq [d]$ is a random set of $\tau$ coordinates. Let $\mathcal{E}_1$ denote this event. The main calculation is to lower bound the probability of this event as follows.

$$\Pr[\mathcal{E}_1] = \frac{1}{n^d} \sum_{\mathbf{y} \in \boldsymbol{B}} \Pr[R \cap C_{\mathbf{y}} \neq \emptyset] \ge \frac{1}{n^d} \sum_{\mathbf{y} \in \boldsymbol{B}} \left[1 - \left(1 - \frac{|C_{\mathbf{y}}|}{d}\right)^{\tau}\right] \ge \frac{1}{n^d} \sum_{\mathbf{y} \in \boldsymbol{B}} \left[1 - \exp\left(-\frac{\tau|C_{\mathbf{y}}|}{d}\right)\right]$$

[5] We point out the following minor technicality in our presentation. From Definition 2.1, note that $c_r$ is chosen from $I_r \setminus \{\mathbf{x}_r\}$ and so technically we will never have $c_r = \mathbf{x}_r$. However, note that Step 4 of Alg. 1 also runs the upward path + downward shift tester using walk length $\tau - 1$ and this is equivalent to setting $c_r = \mathbf{x}_r$ in this analysis, so that the first step of the walk is of length 0. Thus, it is sound in this analysis to think of $c_r$ as uniformly chosen from $I_r$ and we make this simplification for ease of reading.

As in Case 1, the RHS can only decrease if we replace $\tau$ with its lower bound (Definition 4.6, (d)) of $\sigma \cdot d^{1/2-O(\delta)}$, and a similar argument as in Case 1 gives

$$\Pr[\mathcal{E}_1] \geq \frac{\sigma}{d^{1/2+O(\delta)}} \cdot \frac{1}{n^d} \sum_{\mathbf{y} \in \boldsymbol{B}} \Phi_H(\mathbf{y}) \underbrace{\geq}_{\text{(Definition 4.6, (c))}} \frac{\varepsilon^2}{d^{1/2+O(\delta)}} \tag{3}$$

As in Case 1, the event $\mathcal{E}_1$ says that the tester has chosen a point $\mathbf{y} \in \boldsymbol{B}$ and there exists $r \in R$ such that there exists an edge $(\mathbf{y} - a\mathbf{e}_r, \mathbf{y}) \in E$ in the subgraph $H$ for some integer $a > 0$. Fix the smallest $r \in R \cap C_{\mathbf{y}}$ and the corresponding edge in $H$. Now define the following good events for the remainder of the tester analysis.

- $\mathcal{E}_2$: Step (2a) chooses $q_r$ satisfying: if $a \leq n/4$, then $2^{q_r} \in [2a, 4a]$; if $a > n/4$, then $2^{q_r} = n$.
- $\mathcal{E}_3$: Step (2b) chooses the interval $I_r \supseteq [\mathbf{y}_r - a, \mathbf{y}_r]$.
- $\mathcal{E}_4$: Step (2c) chooses $c_r$ uniformly[6] from $[\mathbf{y}_r - a, \mathbf{y}_r]$.
- $\mathcal{E}_5$: $f(\mathbf{x}) = 1$.

The final success probability of the tester is at least $\Pr[\wedge_{i=1}^{5} \mathcal{E}_i]$. Firstly, note that $\Pr[\mathcal{E}_2] = \log^{-1} n$ for both cases of the edge length, $a$. Suppose $a \leq n/4$. Then, $\Pr[\mathcal{E}_3 \mid \mathcal{E}_2] \geq 1/2$ by the condition $q_r \geq 2a$ and $\Pr[\mathcal{E}_4 \mid \mathcal{E}_2, \mathcal{E}_3] \geq 1/4$ by the condition $q_r \leq 4a$. If $a > n/4$, then $\Pr[\mathcal{E}_3 \mid \mathcal{E}_2] = 1$, since in this case $I_r = [n]$ and again $\Pr[\mathcal{E}_4 \mid \mathcal{E}_2, \mathcal{E}_3] \geq 1/4$.

Finally, since the edge $(\mathbf{y} - a\mathbf{e}_r, \mathbf{y})$ is blue for walk length $\tau - 1$, by definition (Definition 4.3) we have $\Pr[\mathcal{E}_5 \mid \mathcal{E}_4] \geq 0.01$. Putting everything together, we have

$$\Pr\left[\bigwedge_{i=1}^{5} \mathcal{E}_i\right] \geq \frac{\varepsilon^2}{d^{1/2+O(\delta)}} \cdot \frac{1}{\log n} \cdot \frac{1}{2} \cdot \frac{1}{4} \cdot \frac{1}{100} \geq \frac{\varepsilon^2}{d^{1/2+O(\delta)}}$$

where in the last step we used $n \leq \mathrm{poly}(d)$ and this completes the proof when the nice subgraph is blue. Together, the cases complete the proof of Lemma 4.7.

## 5.2 Tying it Together: Proof of Theorem 2.4

Suppose $f : [n]^d \to \{0, 1\}$ is $\varepsilon$-far from being monotone. Recall the definitions of $\widetilde{I}_f, \widetilde{I}_f^-$ in Definition 3.13. By Claim 3.15, if $\widetilde{I}_f > 9\sqrt{d}$, then $\widetilde{I}_f^- > \sqrt{d}$ and so the tester (Alg. 1) finds a violation in step (2) when $\tau = 1$ with probability $\Omega(d^{-1/2})$. Thus, we will assume $\widetilde{I}_f \leq 9\sqrt{d}$ and so we may invoke Lemma 4.8 which gives us either a nice red subgraph or a nice blue subgraph. Lemma 4.7 then proves that Alg. 1 finds a violating pair and rejects with probability at least $\varepsilon^2 \cdot d^{-(1/2+O(\delta))}$. This proves Theorem 2.4.

[6]The same minor technicality arises here as in the previous subsection. We will never have $c_r = \mathbf{y}_r$ as per Definition 2.1, but Step 3 of Alg. 1 also runs the downward path tester with walk length $\tau - 1$ and this is equivalent to setting $c_r = \mathbf{y}_r$ in this analysis. Thus, it is again sound in this analysis to think of $c_r$ being uniformly chosen from $I_r$.

# 6 Finding Nice Subgraphs

In this section we prove Lemma 4.8 which we restate below.

**Lemma 4.8** (Existence of nice subgraphs)**.** *Let $n, \varepsilon^{-1} \leq \mathrm{poly}(d)$. Suppose $f\colon [n]^d \to \{0,1\}$ is $\varepsilon$-far from monotone and $\widetilde{I}_f \leq 9\sqrt{d}$. Let $\delta > \frac{1}{\log\log nd}$ be a parameter. There exists $0 < \sigma_1 \leq \sigma_2 < 1$, a violation subgraph $H(\boldsymbol{A}, \boldsymbol{B}, E)$, and a power of two $\tau \geq 2$, such that either $H$ is a $(\sigma_1, \tau)$-nice red subgraph or a $(\sigma_2, \tau)$-nice blue subgraph.*

The proof proceeds over multiple steps and constitutes a key technical contribution of the paper. We give a sketch of what is forthcoming.

- In §6.1 we describe the construction of a *seed* regular violation subgraph $G$. This uses the directed isoperimetric result Theorem 3.4 proved in [BCS23] and a "peeling argument" not unlike that present in [KMS18]. At the end of this section, we will fix the parameters $\sigma_1, \sigma_2$ and the walk length $\tau$. In particular, the length $\tau$ will be defined by the *larger* side of this violating bipartite graph.
- In §6.2, we obtain a regular violating graph $H$ that has persistence properties with respect to the walk length $\tau$. In [KMS18] and [BCS23], one obtained this violating graph by simply deleting the non-persistent points from the seed violation subgraph. In our case, since we choose the walk length depending on the larger side, we need to be careful. We use the idea of "translating violation subgraphs" on $G$ (repeatedly) to find a different violation subgraph $H$ with the desired persistence properties.
- In §6.3, we use the graph $H$ to obtain either a nice red subgraph $H_1$ or a nice blue subgraph $H_2$. If most of the edges in $H$ were red, then a simple surgery on $H$ itself gives us $H_1$. On the other hand, if $H$ has few red edges (but has the persistence properties as guaranteed), then we apply the red/blue lemma (Lemma 4.4) to obtain the desired nice blue-subgraph $H_2$. The proof of the red/blue lemma, which is present in §7, uses the translating violation subgraphs idea as well.

Throughout, we assume $f : [n]^d \to \{0,1\}$ is a function which is $\varepsilon$-far from being monotone, $\widetilde{I}_f \leq 9\sqrt{d}$ and $n, \varepsilon^{-1} \leq \mathrm{poly}(d)$. In particular, we fix a constant $c$ so that $nd \leq d^c$. We also fix a $\delta \approx \frac{1}{\log\log nd} = o(1)$.

## 6.1 Peeling Argument to Obtain Seed Regular Violation Subgraph

Recall the definition of the Talagrand objective (Definition 3.5) $\mathrm{Tal}(G)$ of a violation subgraph $G = (\boldsymbol{X}, \boldsymbol{Y}, E)$. Let $G_0$ denote the violation subgraph formed by all violating edges in the fully augmented hypergrid. Theorem 1.4 in [BCS23] (paraphrased in this paper as Theorem 3.4) is that $\mathrm{Tal}(G_0) = \Omega(\varepsilon n^d / \log n)$. Also recall the definitions in Definition 3.2. The following lemma asserts that there exists a subgraph of $G_0$ whose Talagrand objective is not much lower, but satisfies certain regularity properties.

**Lemma 6.1** (Seed Regular Violation Subgraph)**.** *There exists a violation subgraph $G(\boldsymbol{X}, \boldsymbol{Y}, E)$ satisfying the following properties.*

*(a)* $\mathrm{Tal}(G) \geq \varepsilon \cdot d^{-c\delta} \cdot n^d$.

*(b)* $m(G) \geq d^{-3c\delta} \max(|\boldsymbol{X}|\Phi(\boldsymbol{X})\Gamma(\boldsymbol{X}), |\boldsymbol{Y}|\Phi(\boldsymbol{Y})\Gamma(\boldsymbol{Y}))$.

*(c)* *All vertices in $\boldsymbol{X} \cup \boldsymbol{Y}$ are 98-typical.*

*(d)* $|\boldsymbol{X}|, |\boldsymbol{Y}| \geq \frac{\varepsilon}{d^{1/2+c\delta}} \cdot n^d$.

Let us make a few comments before proving the above lemma. Condition (a) shows that the Talagrand objective degrades only by a $d^{o(1)}$ factor. Condition (b) shows that the graph is nearly regular since the RHS term without the $d^{-o(1)}$ term is the maximum value of $m(G)$. This is because $\Phi(\boldsymbol{X})\Gamma(\boldsymbol{X})$ is an upper bound on the maximum degree of any vertex $\mathbf{x} \in \boldsymbol{X}$. Indeed, if one can prove a stronger lemma which replaces the $d^{o(1)}$ terms in (a) and (b) by $\mathrm{polylog}(d)$'s, then the remainder of our analysis could be easily modified to give a $\tilde{O}(\varepsilon^{-2}\sqrt{d})$ tester.

We need a few tools to prove this lemma. Our first claim is a consequence of the subadditivity of the square root function.

**Claim 6.2.** *Consider a partition of (the edges of) a violation subgraph $G$ into $H_1, H_2, \ldots, H_k$. Then $\sum_{j \leq k} \mathrm{Tal}(H_j) \geq \mathrm{Tal}(G)$.*

*Proof.* Let $\chi_j$ denote the coloring of the subgraph $H_j$ that obtains the minimum $\mathrm{Tal}(H_j)$. Since the $H_1, \ldots, H_k$ form a partition, we can aggregate the colors to get a coloring $\chi$ of $G$.

Consider any $\mathbf{z} \in \boldsymbol{X} \cup \boldsymbol{Y}$. Let $\Phi_{H_j,\chi_j}(\mathbf{z})$ be the thresholded degree of $\mathbf{z}$, restricted to the edges colored by $\chi_j$. By the subadditivity of the square root function, $\sum_{j \leq k} \sqrt{\Phi_{H_j,\chi_j}(\mathbf{z})} \geq \sqrt{\sum_{j \leq k} \Phi_{H_j,\chi_j}(\mathbf{z})}$. Observe that thresholded degrees are also subadditive, so $\sum_{j \leq k} \Phi_{H_j,\chi_j}(\mathbf{z}) \geq \Phi_{G,\chi}(\mathbf{z})$. Hence,

$$\sum_{j \leq k} \mathrm{Tal}(H_j) = \sum_{j \leq k} \sum_{\mathbf{z} \in \boldsymbol{X} \cup \boldsymbol{Y}} \sqrt{\Phi_{H_j,\chi_j}(\mathbf{z})} = \sum_{\mathbf{z} \in \boldsymbol{X} \cup \boldsymbol{Y}} \sum_{j \leq k} \sqrt{\Phi_{H_j,\chi_j}(\mathbf{z})} \geq \sum_{\mathbf{z} \in \boldsymbol{X} \cup \boldsymbol{Y}} \sqrt{\Phi_{G,\chi}(\mathbf{z})} \geq \mathrm{Tal}(G) \quad (4)$$

□

**Remark 6.3.** *The proof of Claim 6.2 crucially uses the fact that in the definition of* $\mathrm{Tal}()$, *we minimize over all possible colorings $\chi$'s of the edges. In particular, if we had defined* $\mathrm{Tal}(G)$ *only with respect to the all ones or the all zeros coloring, then the above proof fails. In the remainder of the paper, we will only be using the $\chi \equiv 1$ or $\chi \equiv 0$ colorings, and the curious reader may wonder why we need the definition of* $\mathrm{Tal}(G)$ *to minimize over all colorings. This is exactly the point where we need it. We make this remark because the "uncolored" isoperimetric theorem is much easier to prove than the "colored" version, but the colored/robust version is essential for the tester analysis.*

Our next step is a simple bucketing argument.

**Claim 6.4.** *Consider a violation subgraph $G = (\boldsymbol{X}, \boldsymbol{Y}, E)$. Both of the following are true.*

1. *There exists a subgraph $G' = (\boldsymbol{X}', \boldsymbol{Y}', E')$ of $G$ such that $\mathrm{Tal}(G') \geq \delta^2 \mathrm{Tal}(G)$ and $m(G') \geq (nd)^{-\delta} |\boldsymbol{X}'| \Phi(\boldsymbol{X}') \Gamma(\boldsymbol{X}')$.*
2. *There exists a subgraph $G' = (\boldsymbol{X}', \boldsymbol{Y}', E')$ of $G$ such that $\mathrm{Tal}(G') \geq \delta^2 \mathrm{Tal}(G)$ and $m(G') \geq (nd)^{-\delta} |\boldsymbol{Y}'| \Phi(\boldsymbol{Y}') \Gamma(\boldsymbol{Y}')$.*

*Proof.* We prove item (1) and the proof of item (2) is analogous.

For convenience, we assume that $\delta$ is the reciprocal of a natural number. For each $\mathbf{x} \in \boldsymbol{X}$, we bucket the incident edges as follows. First, for each $a \in [1/\delta]$, let $S_a$ be the set of dimensions $i$, such that the $i$-degree of $\mathbf{x}$ is in the range $[n^{(a-1)\delta}, n^{a\delta})$. Note that $S_1, \ldots, S_{1/\delta}$ forms a partition of the set of coordinates, $[d]$. Now, for each $a, b \in [1/\delta]$, let the $(a, b)$ edge bucket of $\mathbf{x}$, denoted $E_{a,b,\mathbf{x}}$, be defined as follows. If $|S_a| \in [d^{(b-1)\delta}, d^{b\delta})$, then $E_{a,b,\mathbf{x}}$ is the set of all edges incident to $\mathbf{x}$ along dimensions in $S_a$. If $|S_a| \notin [d^{(b-1)\delta}, d^{b\delta})$, then $E_{a,b,\mathbf{x}} = \emptyset$. Observe $\{E_{a,b,\mathbf{x}} \colon a, b \in [1/\delta]\}$ partitions the edges incident to $\mathbf{x}$.

Now, let $G_{a,b}$ denote the subgraph formed by the edge set $\cup_{\mathbf{x}\in\boldsymbol{X}} E_{a,b,\mathbf{x}}$. Let $\boldsymbol{X}_{a,b}$ be the set of vertices in $\boldsymbol{X}$ with non-zero degree in $G_{a,b}$. Observe that $\Phi(\boldsymbol{X}_{a,b}) \leq d^{b\delta}$ and $\Gamma(\boldsymbol{X}_{a,b}) \leq n^{a\delta}$. Moreover, the degree of each $\mathbf{x} \in \boldsymbol{X}_{a,b}$ is at least $d^{(b-1)\delta} \times n^{(a-1)\delta} \geq (nd)^{-\delta}\Phi(\boldsymbol{X}_{a,b})\Gamma(\boldsymbol{X}_{a,b})$. Hence, $m(G_{a,b}) \geq (nd)^{-\delta}|\boldsymbol{X}_{a,b}|\Phi(\boldsymbol{X}_{a,b})\Gamma(\boldsymbol{X}_{a,b})$.

Finally, by construction, the $G_{a,b}$ subgraphs partition the edges of $G$. Hence, by Claim 6.2 we have $\sum_{a,b\in[1/\delta]} \mathrm{Tal}(G_{a,b}) \geq \mathrm{Tal}(G)$. By averaging, there exists some choice of $a, b$ such that $\mathrm{Tal}(G_{a,b}) \geq \delta^2\mathrm{Tal}(G)$. This gives the desired subgraph $G'$. □

Claim 6.4, part 1 above gives the regularity condition only with respect to $\boldsymbol{X}$, and part 2 gives the analogous guarantee with respect to $\boldsymbol{Y}$, but the trouble is in getting both simultaneously. We do an iterative construction using Claim 6.4 to get the simultaneous guarantee.

*Proof.* (part (a) and (b) of Lemma 6.1) By the robust directed Talagrand theorem for hypergrids (Theorem 3.4), there is a violation subgraph $G_0 = (\boldsymbol{X}_0, \boldsymbol{Y}_0, E_0)$ such that $\mathrm{Tal}(G_0) = \Omega(\varepsilon n^d/\log n)$. We construct a series of subgraphs $G_0 \supseteq G_1 \supseteq G_2 \supseteq \cdots \supseteq G_r$ as follows.

Let $i \geq 1$. If $i$ is odd, we apply item (1) of Claim 6.4 to $G_{i-1}$ to get $G_i(\boldsymbol{X}_i, \boldsymbol{Y}_i, E_i)$ with the regularity condition on $\boldsymbol{X}_i$. If $i$ is even, we apply item (2) of Claim 6.4 to $G_{i-1}$ to get $G_i(\boldsymbol{X}_i, \boldsymbol{Y}_i, E_i)$ with the regularity condition on $\boldsymbol{Y}_i$. If $i > 1$ and $m(G_i) \geq (nd)^{-\delta}m(G_{i-1})$, then we terminate the series. By Claim 6.4, the series satisfies the following three properties for all $i \geq 1$.

- $\mathrm{Tal}(G_i) = \Omega(\delta^{2i}\varepsilon n^d/\log n)$.
- If $i$ is odd, $m(G_i) \geq (nd)^{-\delta}|\boldsymbol{X}_i|\Phi(\boldsymbol{X}_i)\Gamma(\boldsymbol{X}_i)$. If $i$ is even, $m(G_i) \geq (nd)^{-\delta}|\boldsymbol{Y}_i|\Phi(\boldsymbol{Y}_i)\Gamma(\boldsymbol{Y}_i)$.
- If the series has not terminated by step $i$, then $m(G_i) < (nd)^{-\delta}m(G_{i-1})$.

The first two statements hold by the guarantees of Claim 6.4 and the fact that $\mathrm{Tal}(G_0) = \Omega(\varepsilon n^d/\log n)$. The third statement holds simply by the termination condition for the sequence. The trivial bound on the number of edges is $m(G_0) \leq nd \cdot n^d$. The third bullet point yields $m(G_i) < (nd)^{-i\delta} \cdot nd \cdot n^d$, if the series has not terminated by step $i$.

**Claim 6.5.** *The series terminates in at most $3/\delta$ steps.*

*Proof.* Suppose not. Noting that $m(G_i) \geq \mathrm{Tal}(G_i)$ (Observation 3.6), we get the following chain of inequalities using the properties of our subgraph graph $G_{3/\delta}$.

$$(nd)^{-(3/\delta)\cdot\delta} \cdot nd \cdot n^d > m(G_i) \geq \mathrm{Tal}(G_i) = \Omega(\delta^{6/\delta}\varepsilon n^d/\log n) \quad\Longrightarrow\quad (nd)^{-2} = \Omega(\delta^{6/\delta}\varepsilon/\log n)$$

Note that we may assume $\varepsilon \geq 1/d$ and so $C\varepsilon/\log n \geq (nd)^{-1}$ for any constant $C$. Thus we have $(nd)^{-1} \geq \delta^{6/\delta}$. Given that $\delta > 1/\log\log nd$, this inequality is a contradiction. □

By the previous claim the series terminates in some $r \leq 3/\delta$ steps, producing $G_r(\boldsymbol{X}_r, \boldsymbol{Y}_r, E_r)$, which we claim has the desired properties to prove conditions (a) and (b) of Lemma 6.1. Since $r \leq 3/\delta$, $\mathrm{Tal}(G_r) = \Omega(\delta^{6/\delta}\varepsilon n^d/\log n)$. Note that since $\delta > 1/\log\log nd$, we have

$$\delta^{6/\delta} > (\log\log nd)^{-\frac{6}{\delta}} = (nd)^{-\frac{6}{\delta}\cdot\frac{\log\log\log nd}{\log nd}} > (nd)^{-\delta^2} > (nd)^{-\delta}\cdot\log n > d^{-c\delta}\log n$$

where the second to last step holds because $\frac{6 \log\log\log nd}{\log d} \ll \left(\frac{1}{\log\log nd}\right)^3 < \delta^3$. The last inequality used $nd \leq d^c$. This proves condition (a). Towards proving condition (b), note that $C\delta^{6/\delta}/\log n \geq (nd)^{-\delta}$ for any constant $C$.

Let's assume without loss of generality that $r$ is even. Thus we have $m(G_r) \geq (nd)^{-\delta}|\boldsymbol{Y}_r|\Phi(\boldsymbol{Y}_r)\Gamma(\boldsymbol{Y}_r)$ by the second bullet point above. Next, since the series terminated at step $r$, we have

$$m(G_r) \geq (nd)^{-\delta} m(G_{r-1}) \geq (nd)^{-2\delta}|\boldsymbol{X}_{r-1}|\Phi(\boldsymbol{X}_{r-1})\Gamma(\boldsymbol{X}_{r-1}) \geq (nd)^{-2\delta}|\boldsymbol{X}_r|\Phi(\boldsymbol{X}_r)\Gamma(\boldsymbol{X}_r)$$

where the second inequality is again by the second bullet point above and the fact that $i-1$ is odd and the third inequality is simply because $G_r$ is a subgraph of $G_{r-1}$. Again using $nd \leq d^c$, we have $(nd)^{-\delta} \geq d^{-c\delta}$ and so we get that $G_r$ satisfies conditions (a) and (b) of Lemma 6.1. □

*Proof.* (Conditions (c) and (d) Lemma 6.1) To obtain condition (c), we simply remove the non-typical points. Recall the definition of $c$-typical points (Definition 3.20). By Claim 3.21, the number of points that are not 98-typical is at most $(\varepsilon/d)^{93}n^d$. Thus, removing all such vertices can decrease $\mathrm{Tal}(G)$ by at most $(\varepsilon/d)^{93}n^d \cdot \sqrt{d}$ which is negligible compared to the RHS in condition (a). Thus, we remove all such vertices from $G$ and henceforth assume that all of $\boldsymbol{X} \cup \boldsymbol{Y}$ is 98-typical.

Condition (d) follows from condition (a). Consider the constant coloring $\chi \equiv 1$ and observe that

$$|\boldsymbol{X}|\sqrt{d} \geq \mathrm{Tal}_{\chi\equiv 1}(G) \geq \mathrm{Tal}(G) \geq \varepsilon \cdot d^{-c\delta} \cdot n^d.$$

where the first inequality follows from the trivial observation that the maximum $\Phi_G(\mathbf{x})$ can be is $d$. Using the coloring $\chi \equiv 0$ proves the same lower bound for $|\boldsymbol{Y}|$. □

### 6.1.1 Choice of the walk length

We end this section by specifying what the parameters $\sigma_1, \sigma_2$ and $\tau$ are going to be in Lemma 4.8. We now make the assumption $|\boldsymbol{X}| \leq |\boldsymbol{Y}|$. Given Remark 2.3, this is without loss of generality; this fact would be true either in $f$ or in $g$, and running steps 2, 3, 5 on $f$ is equivalent to running steps 2, 3, 4 on $g$. The violation subgraphs for $f$ and $g$ are isomorphic. Then,

$$\sigma_1 = \sigma_{\boldsymbol{X}} := \frac{|\boldsymbol{X}|}{n^d} \quad \text{and} \quad \sigma_2 = \sigma_{\boldsymbol{Y}} := \frac{|\boldsymbol{Y}|}{n^d}$$

and set $\tau$ to be the unique power of two such that

$$\frac{1}{2}\lceil \sigma_{\boldsymbol{Y}} \cdot d^{1/2-7c\delta} \rceil < \tau - 1 \leq \lceil \sigma_{\boldsymbol{Y}} \cdot d^{1/2-7c\delta} \rceil.$$

We conclude the subsection by establishing the following upper bound on the number of vertices which are not up-persistent.

**Claim 6.6.** *We may assume that*

- *the number of vertices* $\mathbf{x} \in \boldsymbol{X}$ *where* $f(\mathbf{x}) = 1$ *that are not* $(\tau - 1, \log^{-5} d)$*-up-persistent is at most* $d^{-6c\delta} \cdot |\boldsymbol{X}|$*, and*
- *the number of vertices* $\mathbf{y} \in \boldsymbol{Y}$ *where* $f(\mathbf{y}) = 0$ *that are not* $(\tau - 1, \log^{-5} d)$*-up-persistent is at most* $d^{-6c\delta} \cdot |\boldsymbol{Y}|$.

*Proof.* The statement for points where $f(\mathbf{x}) = 1$ is implied by item (4) of Lemma 6.1, for otherwise the tester succeeds with the desired probability when it runs the upward path tester with walk length $\tau - 1$ (step (2) of Alg. 1).

Now we prove the statement for points where $f(\mathbf{y}) = 0$. By Claim 3.17, the total number of $(\tau - 1, \log^{-5} d)$-non-persistent vertices is at most $C_{per}\tau \cdot \log^5 d \cdot \frac{1}{\sqrt{d}} \cdot n^d \leq \sigma_{\boldsymbol{Y}} \cdot d^{-6c\delta} \cdot n^d$, where we have simply used $\log^5 d \ll d^{c\delta}$ and our definition of $\tau$. $\square$

## 6.2 Using 'Persist-or-Blow-up' Lemma to obtain Down-Persistence

Lemma 6.1 provides a seed violation subgraph which has a large Talagrand objective and has regularity properties. Claim 6.6 shows that we may assume these vertices are up-persistent with respect to walk length of $\tau - 1$. However, we may not have down persistence. In particular, it could be $|\boldsymbol{X}| \ll |\boldsymbol{Y}|$ and if we try to apply Claim 3.17 and remove all nodes from $\boldsymbol{X}$ which are not $(\tau - 1, 0.6)$-down-persistent, we may end up removing everything. To obtain a subgraph with down-persistence properties, we need to apply a translation procedure which is encapsulated in the lemma below. The proof of the lemma is deferred to §8.

**Lemma 6.7** (Persist-or-Blow-up Lemma). *Consider a violation subgraph $G = (\boldsymbol{X}, \boldsymbol{Y}, E)$ such that all vertices in $G$ are $c$-typical where $c \leq 99$ and $(\ell, \log^{-5} d)$-up persistent where $1 \leq \ell \leq \sqrt{d}/\log^5(d/\varepsilon)$. Then, there exists a violation subgraph $G' = (\boldsymbol{X}', \boldsymbol{Y}', E')$ where all vertices are $(c + \frac{\ell}{\sqrt{d}})$-typical and satisfying one of the following conditions.*

1. *Down-persistent case:*
   (a) *All vertices in $\boldsymbol{X}'$ are $(\ell, 0.6)$-down persistent.*
   (b) $m(G') \geq m(G)/\log^5 d$.
   (c) $D(\boldsymbol{X}') \leq D(\boldsymbol{X})$, *and* $\forall i \in [d], \Gamma_i(\boldsymbol{X}') \leq \Gamma_i(\boldsymbol{X})$
   (d) $D(\boldsymbol{Y}') \leq D(\boldsymbol{Y})$, *and* $\forall i \in [d]$, $\Gamma_i(\boldsymbol{Y}') \leq \Gamma_i(\boldsymbol{Y})$.
2. *Blow-up case:*
   (a) $m(G') \geq 2(1 - 3\log^{-3} d) \cdot m(G)$.
   (b) $D(\boldsymbol{X}') \leq D(\boldsymbol{X})$, *and* $\forall i \in [d], \Gamma_i(\boldsymbol{X}') \leq \Gamma_i(\boldsymbol{X})$
   (c) $D(\boldsymbol{Y}') \leq 2D(\boldsymbol{Y})$, *and* $\forall i \in [d]$, $\Gamma_i(\boldsymbol{Y}') \leq 2\Gamma_i(\boldsymbol{Y})$.

That is, the application of the above lemma either gives the violation subgraph we need, or it gives us a violation subgraph with around double the edges. In the remainder of this section we use Lemma 6.7 and the graph $G(\boldsymbol{X}, \boldsymbol{Y}, E)$ derived in the previous section to prove the following lemma.

**Lemma 6.8** (Down-Persistent Violation Subgraph). *Let $G(\boldsymbol{X}, \boldsymbol{Y}, E)$ be the subgraph asserted in Lemma 6.1. There exists a natural number $s \leq \log^3 d$ and a violation subgraph $H(\boldsymbol{A}, \boldsymbol{B}, E)$ with the following properties.*

1. $m(H) \geq 2^s \frac{m(G)}{\log^7 d}$.
2. $\Gamma(\boldsymbol{A}) \leq \Gamma(\boldsymbol{X})$ *and* $\Gamma(\boldsymbol{B}) \leq 2^s \Gamma(\boldsymbol{Y})$.
3. $D(\boldsymbol{A}) \leq D(\boldsymbol{X})$ *and* $D(\boldsymbol{B}) \leq 2^s D(\boldsymbol{Y})$.

4. *All vertices in $\boldsymbol{A} \cup \boldsymbol{B}$ are $(\tau - 1, \log^{-5} d)$-up-persistent and $99$-typical.*

5. *All vertices in $\boldsymbol{A}$ are $(\tau - 1, 0.6)$-down-persistent.*

*Proof.* We use Lemma 6.7 to define the following process generating a sequence of violation subgraphs. The initial graph is $G_0 = (\boldsymbol{X}_0, \boldsymbol{Y}_0, E_0)$ which is the seed regular violation subgraph obtained from Lemma 6.1.

For each $i \geq 1$:

1. Obtain $G'_{i-1}$ by removing all vertices from $\boldsymbol{X}_{i-1} \cup \boldsymbol{Y}_{i-1}$ that are not $(\tau - 1, \log^{-5} d)$-up-persistent.
2. Invoke Lemma 6.7 with walk length $\tau - 1$ on $G'_{i-1}$ to obtain $G_i = (\boldsymbol{X}_i, \boldsymbol{Y}_i, E_i)$.
3. If $G_i$ satisfies the down persistence condition of Lemma 6.7 then halt and return $G_i$.
4. If $G_i$ satisfies the blowup condition of Lemma 6.7, then continue.

By Lemma 6.7, if the process does not halt on step $i$, then we have the following recurrences.

- $m(G_i) \geq 2(1 - 3\log^{-3} d) \cdot m(G'_{i-1})$.
- $D(\boldsymbol{X}_i) \leq D(\boldsymbol{X}_{i-1})$, $\Gamma(\boldsymbol{X}_i) \leq \Gamma(\boldsymbol{X}_{i-1})$, $D(\boldsymbol{Y}_i) \leq 2D(\boldsymbol{Y}_{i-1})$, $\Gamma(\boldsymbol{Y}_i) \leq 2\Gamma(\boldsymbol{Y}_{i-1})$.

Furthermore, we have the following claim that bounds the number of edges lost in step (1).

**Claim 6.9.** *For every $i \geq 1$, we have $m(G'_{i-1}) \geq m(G_{i-1}) - d^{-2c\delta} \cdot 2^{i-1} \cdot m(G)$.*

*Proof.* By Claim 6.6, the number of vertices we remove from $\boldsymbol{X}_{i-1}$ in step (1) is at most $d^{-6c\delta} \cdot |\boldsymbol{X}|$ and the number of vertices we remove from $\boldsymbol{Y}_{i-1}$ in step (1) is at most $d^{-6c\delta} \cdot |\boldsymbol{Y}|$. The number of edges we remove by deleting these vertices from $\boldsymbol{Y}_{i-1}$ is at most

$$d^{-6c\delta}|\boldsymbol{Y}|D(\boldsymbol{Y}_{i-1}) \leq d^{-6c\delta}2^{i-1}|\boldsymbol{Y}|D(\boldsymbol{Y}) \leq d^{-3c\delta}2^{i-1}m(G) \tag{5}$$

where in the second inequality we used $D(\boldsymbol{Y}) \leq \Phi(\boldsymbol{Y})\Gamma(\boldsymbol{Y})$ and the regularity property on $G$ (item (2) of Lemma 6.1).

An analogous argument bounds the number of removed edges when we delete non-persistent vertices from $\boldsymbol{X}_{i-1}$. Thus the total number of edges removed is at most $d^{-2c\delta}2^{i-1}m(G)$. $\square$

**Claim 6.10.** *If $i \leq \log^3 d$ and the process has not halted by step $i$, then $m(G_i) \geq \Omega(2^i m(G))$.*

*Proof.* For brevity, let $\alpha = 2(1 - 3\log^{-3} d)$ and $\beta = d^{-2c\delta}m(G)$. Using the above bounds, we get the recurrence

$$m(G_i) \geq \alpha \cdot m(G'_{i-1}) \geq \alpha(m(G_{i-1}) - \beta 2^{i-1}).$$

Expanding this recurrence yields $m(G_i) \geq \alpha^i m(G) - \beta \sum_{j=1}^{i} \alpha^j \cdot 2^{i-j}$. Observe that the subtracted term can be bounded as

$$\beta \sum_{j=1}^{i} \alpha^j \cdot 2^{i-j} = d^{-2c\delta}2^i m(G) \sum_{j=1}^{i} (1 - 3\log^{-3} d)^j \leq d^{-c\delta}2^i m(G)$$

simply using the fact that $i \le \log^3 d \ll d^{c\delta}$. The first term is

$$\alpha^i m(G) = 2^i(1 - 3\log^{-3} d)^i m(G) \ge C \cdot 2^i m(G)$$

for some constant $C$. Combining the above two bounds completes the proof. $\square$

**Claim 6.11.** *The above process halts in $s \le \log^3 d$ iterations.*

*Proof.* Suppose that the above process has not halted by step $i = \log^3 d$. By the previous claim, the number of edges in $G_i$ is at least $C \cdot 2^i m(G) = C \cdot d^{\log^2 d} m(G)$ for some constant $C$. By Observation 3.6, note that $m(G) \ge \mathrm{Tal}(G)$ and thus is $\ge \varepsilon \cdot d^{-c\delta} \cdot n^d$ by item (1) of Lemma 6.1. Thus, the number of edges in $G_i$ is at least $C \cdot \varepsilon \cdot d^{\log^2 d - c\delta} n^d$. Note that the total number of edges in the fully augmented hypergrid is at most $nd \cdot n^d$. Moreover, recall that we are assuming $nd \le d^c$ and $\varepsilon \ge d^{-1/2}$. Therefore, $m(G_i) \gg nd \cdot n^d$ and this is a contradiction. $\square$

By Claim 6.11 and Lemma 6.7, the process halts in some $s \le \log^3 d$ number of steps producing $G_s(\boldsymbol{X}_s, \boldsymbol{Y}_s, E_s)$ with the following properties.

- $m(G_s) \ge 2^s \cdot \frac{m(G)}{\log^6 d}$.
- All vertices in $\boldsymbol{X}_s$ are $(\tau - 1, 0.6)$-down-persistent.
- $\Gamma(\boldsymbol{X}_s) \le \Gamma(\boldsymbol{X})$ and $\Gamma(\boldsymbol{Y}_s) \le 2^s \Gamma(\boldsymbol{Y})$.
- $D(\boldsymbol{X}_s) \le D(\boldsymbol{X})$ and $D(\boldsymbol{Y}_s) \le 2^s D(\boldsymbol{Y})$.

Note that by Lemma 6.7 and (c) of Lemma 6.1, all vertices in $G_1, \ldots, G_s$ are $(98 + \frac{s\tau}{\sqrt{d}})$-typical. Moreover, by our choice of $\tau$, we have $s\tau \ll \sqrt{d}$ and so all vertices in $G_1, \ldots, G_s$ are 99-typical.

One last time, we remove all vertices in $\boldsymbol{X}_s \cup \boldsymbol{Y}_s$ that are not $(\tau - 1, \log^{-5} d)$-up-persistent and obtain our final graph $H(\boldsymbol{A}, \boldsymbol{B}, E)$. Using a similar argument made above in (5), the number of edges that are removed by deleting the non-persistent vertices from $\boldsymbol{Y}_s$ is at most

$$d^{-6c\delta}|\boldsymbol{Y}|D(\boldsymbol{Y}_s) \le 2^s d^{-6c\delta}|\boldsymbol{Y}|D(\boldsymbol{Y}) \le 2^s d^{-3c\delta} m(G) \le d^{-3c\delta} m(G_s) \log^6 d \le d^{-2c\delta} m(G_s)$$

and an analogous argument bounds the number of edges lost when we remove the non-persistent vertices from $\boldsymbol{X}_s$. Thus we have $m(H) \ge m(G_s)(1 - d^{-c\delta}) \ge 2^s \frac{m(G)}{\log^7 d}$ and this completes the proof of Lemma 6.8. $\square$

### 6.3 Using Red/Blue Lemma to Obtain the Final Red or Blue Nice Subgraph

In this section, we prove Lemma 4.8 using the violation subgraph $H(\boldsymbol{A}, \boldsymbol{B}, E)$ obtained in the previous section (Lemma 6.8) and the red/blue lemma, Lemma 4.4. We split into two cases depending on how many edges in $H$ are red.

### 6.3.1 Case 1: At least half the edges of $H$ are red

In this case, we consider the graph $H_1$ by simply removing all the non-red edges. We claim that $H_1$ makes progress towards a $(\sigma_1, \tau)$-nice red subgraph (Definition 4.5). Condition (a) holds by definition. Condition (b) is satisfied due to Lemma 6.8, condition 5). Condition (e) is satisfied because $\tau - 1 \geq 0.5\sigma_{\boldsymbol{Y}} d^{0.5-7c\delta}$ and $\sigma_{\boldsymbol{Y}} \geq \sigma_{\boldsymbol{X}} = \sigma_1$. We need to establish condition (c) and (d). That is, we need to establish

(c) $\sigma_{\boldsymbol{X}} \cdot \Phi_H(\mathbf{x}) \leq \sqrt{d}$ for all $\mathbf{x} \in \boldsymbol{A}$

(d) $\sigma_{\boldsymbol{X}} \sum_{\mathbf{x} \in \boldsymbol{A}} \Phi_H(\mathbf{x}) \geq \varepsilon^2 \cdot n^d \cdot d^{-6c\delta}$

Let $\boldsymbol{A}' \subseteq \boldsymbol{A}$ be the vertices $\mathbf{x} \in \boldsymbol{A}$ which have $\Phi_H(\mathbf{x}) > \frac{\sqrt{d}}{\sigma_{\boldsymbol{X}}}$. If $|\boldsymbol{A}'| \geq d^{-5c\delta}|\boldsymbol{X}|$, then simply consider $H_1(\boldsymbol{A}', \boldsymbol{B}, E')$ by deleting all vertices not in $\boldsymbol{A}'$ from $\boldsymbol{A}$. Conditions (a), (b), (e) still hold, and (c) holds by design. Furthermore,

$$\sum_{\mathbf{x} \in \boldsymbol{A}'} \Phi_{H_1}(\mathbf{x}) \geq d^{-5c\delta}|\boldsymbol{X}| \cdot \frac{\sqrt{d}}{\sigma_{\boldsymbol{X}}} \;\Rightarrow\; \sigma_{\boldsymbol{X}} \sum_{\mathbf{x} \in \boldsymbol{A}'} \Phi_{H_1}(\mathbf{x}) \geq d^{-5c\delta} \cdot \frac{\varepsilon}{d^{1/2+c\delta}} \cdot n^d \cdot \sqrt{d} = \varepsilon \cdot n^d \cdot d^{-6c\delta}$$

where we used Lemma 6.1, part (d) for the lower bound on $|\boldsymbol{X}|$. Note that this implies something slightly stronger than condition (d) above (the exponent of $\varepsilon$ is 1).

Therefore, we may assume $|\boldsymbol{A}'| \leq d^{-5c\delta}|\boldsymbol{X}|$. In this case, let $H_1 = (\boldsymbol{A} \setminus \boldsymbol{A}', \boldsymbol{B}, E')$ where we simply remove the $\boldsymbol{A}'$ vertices. The number of edges this destroys is at most

$$d^{-5c\delta} D(\boldsymbol{A})|\boldsymbol{X}| \leq d^{-5c\delta} D(\boldsymbol{X})|\boldsymbol{X}| \leq d^{-2c\delta} m(G) \leq d^{-c\delta} m(H)$$

where in the second inequality we used $D(\boldsymbol{X}) \leq \Phi(\boldsymbol{X})\Gamma(\boldsymbol{X})$ and the regularity property (Lemma 6.1, property (b)) of $G$. Thus, the number of edges we've discarded is negligible, and condition (c) holds. In particular, the number of edges in $H_1$ is at least $m(H)/2$. We now prove condition (d) also holds.

**Claim 6.12.** $\sigma_{\boldsymbol{X}} \sum_{\mathbf{x} \in \boldsymbol{A} \setminus \boldsymbol{A}'} \Phi_{H_1}(\mathbf{x}) \geq \varepsilon^2 \cdot n^d \cdot d^{-6c\delta}$.

*Proof.* For any $\mathbf{x} \in \boldsymbol{A} \setminus \boldsymbol{A}'$, we have $\Phi_{H_1}(\mathbf{x}) \geq \frac{D(\mathbf{x})}{\Gamma(\mathbf{x})}$ and thus $\sum_{\mathbf{x} \in \boldsymbol{A} \setminus \boldsymbol{A}'} \Phi_H(\mathbf{x}) \geq \frac{m(H)/2}{\Gamma(\boldsymbol{A})}$. Since $\Gamma(\boldsymbol{A}) \leq \Gamma(\boldsymbol{X})$ we have

$$\begin{aligned}\sum_{\mathbf{x} \in \boldsymbol{A} \setminus \boldsymbol{A}'} \Phi_H(\mathbf{x}) &\geq \frac{m(H)}{2\Gamma(\boldsymbol{A})} \geq \frac{2^s \cdot m(G)}{2\Gamma(\boldsymbol{X}) \log^7 d} \geq \frac{d^{-3c\delta}|\boldsymbol{X}|\Phi(\boldsymbol{X})\Gamma(\boldsymbol{X})}{2\Gamma(\boldsymbol{X}) \log^7 d} \\ &\geq d^{-4c\delta}|\boldsymbol{X}|\Phi(\boldsymbol{X}) \geq d^{-4c\delta} \sum_{\mathbf{x} \in \boldsymbol{X}} \Phi_G(\mathbf{x}). \end{aligned} \tag{6}$$

where in the second inequality we used (P1) to lower bound the number of edges in $H$ with that of $G$. In the third inequality we used the regularity property (property (b) of Lemma 6.1), in the fourth we used $d^{c\delta} \gg 2\log^7 d$ for large enough $d$, and the fifth inequality uses the trivial upper bound $\Phi(\boldsymbol{X}) \geq \Phi_G(\mathbf{x})$ for all $\mathbf{x} \in \boldsymbol{X}$.

Now we apply the fact (Lemma 6.1, condition 1) that $\mathrm{Tal}(G)$ is large. Using the coloring $\chi \equiv 1$ for edges in $G$, we get

$$\sum_{\mathbf{x} \in \boldsymbol{X}} \sqrt{\Phi_G(\mathbf{x})} \geq \mathrm{Tal}(G) \geq \varepsilon \cdot d^{-c\delta} \cdot n^d \;\Rightarrow\; \mathbb{E}_{\mathbf{x} \in \boldsymbol{X}}[\sqrt{\Phi_G(\mathbf{x})}] \geq \frac{\varepsilon \cdot d^{-c\delta}}{\sigma_{\boldsymbol{X}}}$$

Jensen's inequality gives

$$\mathbb{E}_{\mathbf{x}\in\boldsymbol{X}}[\Phi_G(\mathbf{x})] \geq \frac{\varepsilon^2 \cdot d^{-2c\delta}}{\sigma_{\boldsymbol{X}}^2} \quad\Rightarrow\quad \frac{\sigma_{\boldsymbol{X}}^2}{|\boldsymbol{X}|}\sum_{\mathbf{x}\in\boldsymbol{X}}\Phi_G(\mathbf{x}) \geq \varepsilon^2 d^{-2c\delta} \quad\Rightarrow \sigma_{\boldsymbol{X}}\sum_{\mathbf{x}\in\boldsymbol{X}}\Phi_G(\mathbf{x}) \geq \varepsilon^2 d^{-2c\delta} n^d$$

Plugging into (6) proves the claim. □

#### 6.3.2 Case 2: At most half the edges of $H$ are red

In this case we invoke the Red/Blue lemma, Lemma 4.4 to obtain a violation subgraph $H_2 = (\boldsymbol{L}, \boldsymbol{R}, E')$ with the following key properties.

(P1) All edges are blue and $m(H) \geq 2^s \frac{m(G)}{7\log^7 d}$.

(P2) $\Gamma(\boldsymbol{R}) \leq \Gamma(\boldsymbol{B}) \leq 2^s \cdot \Gamma(\boldsymbol{Y})$.

(P3) $D(\boldsymbol{R}) \leq D(\boldsymbol{B}) \leq 2^s \cdot D(\boldsymbol{Y})$.

We claim that $H_s$ makes progress towards a $(\sigma_2, \tau)$-nice blue subgraph (Definition 4.6). Condition (a) holds by definition. Condition (d) is satisfied because $\tau \geq 0.5\sigma_{\boldsymbol{Y}} d^{0.5-7c\delta}$ and $\sigma_{\boldsymbol{Y}} = \sigma_2$. We need to establish condition (b) and (c). That is, we need to establish

(b) $\sigma_{\boldsymbol{Y}} \cdot \Phi_H(\mathbf{y}) \leq \sqrt{d}$ for all $\mathbf{x} \in \boldsymbol{R}$

(c) $\sigma_{\boldsymbol{Y}} \sum_{\mathbf{y}\in\boldsymbol{R}} \Phi_H(\mathbf{y}) \geq \varepsilon^2 \cdot n^d \cdot d^{-6c\delta}$

As in Case 1, we begin by removing low degree vertices. Let $\boldsymbol{R}' \subseteq \boldsymbol{R}$ be the vertices $\mathbf{y} \in \boldsymbol{R}$ which have $\Phi_H(\mathbf{y}) > \frac{\sqrt{d}}{\sigma_{\boldsymbol{Y}}}$. If $|\boldsymbol{R}'| \geq d^{-5c\delta}|\boldsymbol{Y}|$, then we would just focus on $H_2(\boldsymbol{R}', \boldsymbol{L}, E')$ and this would satisfy (b) and (c) for a very similar reason as in Case 1. And so, we may assume $|\boldsymbol{R}'|$ is smaller than $d^{-5c\delta}|\boldsymbol{Y}|$ and we define $H_2(\boldsymbol{L}, \boldsymbol{R} \setminus \boldsymbol{R}', E')$, and this leads to a negligible decrease in the number of edges. Condition (b) holds by design, and the proof that condition (c) holds is similar. We provide it for completeness.

**Claim 6.13.** $\sigma_{\boldsymbol{Y}} \sum_{\mathbf{y}\in\boldsymbol{R}\setminus\boldsymbol{R}'} \Phi_{H_2}(\mathbf{y}) \geq \varepsilon^2 \cdot n^d \cdot d^{-6c\delta}$.

*Proof.* For any $\mathbf{y} \in \boldsymbol{R} \setminus \boldsymbol{R}'$, we have $\Phi_{H_2}(\mathbf{y}) \geq \frac{D(\mathbf{y})}{\Gamma(\mathbf{y})}$ and thus $\sum_{\mathbf{y}\in\boldsymbol{R}\setminus\boldsymbol{R}'} \Phi_{H_2}(\mathbf{y}) \geq \frac{m(H)/2}{\Gamma(\boldsymbol{R})}$. Since $\Gamma(\boldsymbol{R}) \leq 2^s \cdot \Gamma(\boldsymbol{Y})$ we have

$$\begin{aligned}\sum_{\mathbf{y}\in\boldsymbol{R}\setminus\boldsymbol{R}'} \Phi_H(\mathbf{y}) &\geq \frac{m(H)}{2\Gamma(\boldsymbol{R})} \geq \frac{2^s \cdot m(G)}{2^s \cdot 14\Gamma(\boldsymbol{Y})\log^7 d} \geq \frac{d^{-3c\delta}|\boldsymbol{Y}|\Phi(\boldsymbol{Y})\Gamma(\boldsymbol{Y})}{14\Gamma(\boldsymbol{Y})\log^7 d} \\ &\geq d^{-4c\delta}|\boldsymbol{Y}|\Phi(\boldsymbol{Y}) \geq d^{-4c\delta}\sum_{\mathbf{y}\in\boldsymbol{Y}}\Phi_G(\mathbf{y}). \end{aligned} \tag{7}$$

where in the second inequality we used Lemma 6.8, part 1, to lower bound the number of edges in $H$ with that of $G$, the original seed graph from Lemma 6.1. In the third inequality we used the regularity property (property 2 of Lemma 6.1), in the fourth we used $d^{c\delta} \gg 14\log^7 d$ for large enough $d$, and the fifth inequality uses the trivial upper bound $\Phi(\boldsymbol{Y}) \geq \Phi_G(\mathbf{y})$ for all $\mathbf{y} \in \boldsymbol{Y}$.

The rest of the proof is the same as Case 1 except we apply the coloring $\chi \equiv 0$ for edges in $G$. We omit this very similar calculation. □

These two cases conclude the proof of Lemma 4.8. All that remains is to prove the Red/Blue lemma, Lemma 4.4 and the Persist-or-Blow-up lemma, Lemma 6.7. We prove these in the subsequent two sections, and both of these use the translation of violation subgraphs idea.

# 7 Proof of the Red/Blue Lemma, Lemma 4.4

Let us recall the red/blue lemma.

**Lemma 4.4** (Red/Blue Lemma). *Let $G(\boldsymbol{X}, \boldsymbol{Y}, E)$ be a violation subgraph and $1 \leq \ell \leq \sqrt{d}/\log^5(d/\varepsilon)$ be a walk length such that the following hold.*

1. *At most half the edges are red for walk length $\ell$.*
2. *All vertices in $\boldsymbol{X} \cup \boldsymbol{Y}$ are $(\ell, \log^{-5} d)$-up-persistent.*
3. *All vertices in $\boldsymbol{X} \cup \boldsymbol{Y}$ are 99-typical.*

*Then there exists another violation subgraph $H(\boldsymbol{L}, \boldsymbol{R}, E')$ such that*

1. *All edges are blue for walk length $\ell$ and $m(H) \geq m(G)/7$.*
2. *$\Gamma(\boldsymbol{L}) \leq \Gamma(\boldsymbol{X})$ and $\Gamma(\boldsymbol{R}) \leq \Gamma(\boldsymbol{Y})$.*
3. *$D(\boldsymbol{L}) \leq D(\boldsymbol{X})$ and $D(\boldsymbol{R}) \leq D(\boldsymbol{Y})$.*

*Proof.* We first recall the definition of $p_{\mathbf{x},\ell}(\mathbf{x}')$ in Definition 3.23. For a fixed $\mathbf{x}$, consider the process of sampling a hypercube $\boldsymbol{H} \sim \mathbb{H}(\mathbf{x})$ and then sampling $\mathbf{z} \sim \mathcal{U}_{\boldsymbol{H},\ell}(\mathbf{x})$. Recall from Fact 3.11 that this is one of three equivalent ways of expressing our random walk distribution. Given $\mathbf{x}, \mathbf{x}', \ell$, we have

$$p_{\mathbf{x},\ell}(\mathbf{x}') = \Pr\left[\mathbf{x}, \mathbf{x}' \in \boldsymbol{H}_{100} \text{ and } \mathbf{z} = \mathbf{x}'\right].$$

We use these values to set up a flow problem as follows.

Recall the definition of red and blue edges (Definition 4.2 and Definition 4.3). Let $B$ denote the set of all edges in the fully augmented hypergrid that are blue for walk length $\ell$. For every non-red edge $(\mathbf{x}, \mathbf{y})$ of $G$ and every shift $\mathbf{s} \in \operatorname{supp}(\mathcal{US}_\ell(\mathbf{x}))$, if the edge $e = (\mathbf{x}+\mathbf{s}, \mathbf{y}+\mathbf{s})$ is blue, then we put $p_{\mathbf{x},\ell}(\mathbf{x}+\mathbf{s})$ units of flow on $e$.

**Claim 7.1.** *Every non-red edge of $G$ inserts at least $0.95$ units of flow in $B$.*

*Proof.* Fix a non-red edge $(\mathbf{x}, \mathbf{y})$, and let $i$ denote its dimension. Generate $\boldsymbol{H} \sim \mathbb{H}(\mathbf{x})$ and $\mathbf{s} \sim \mathcal{US}_{\boldsymbol{H},\ell}(\mathbf{x})$. Note that it is equivalent to directly sample $\mathbf{s} \sim \mathcal{US}_\ell(\mathbf{x})$. We then consider the random edge $e = (\mathbf{x}+\mathbf{s}, \mathbf{y}+\mathbf{s})$. We set $\mathbf{x}' = \mathbf{x}+\mathbf{s}$ and $\mathbf{y}' = \mathbf{y}+\mathbf{s}$. Let us define the following series of events. (i) $\mathcal{E}_1$: $\mathbf{s}_i = 0$. (ii) $\mathcal{E}_2$: $f(\mathbf{x}') = 1$. (iii) $\mathcal{E}_3$: $f(\mathbf{y}') = 0$. (iv) $\mathcal{E}_4$: at least half of $I(\mathbf{x}', \mathbf{y}')$ is *not* $\ell$-mostly-zero-below, (v) $\mathcal{E}_5$: $\mathbf{x}, \mathbf{x}' \in \boldsymbol{H}_{100}$. We will show that whenever $\mathcal{E}_2$, $\mathcal{E}_3$, and $\mathcal{E}_4$ occur, the edge $(\mathbf{x}', \mathbf{y}')$ is blue by definition. Therefore, recalling the definition of $p_{\mathbf{x},\ell}(\mathbf{x}')$, the edge $(\mathbf{x}, \mathbf{y})$ inserts at least $\Pr[\wedge_{j=1}^5 \mathcal{E}_j]$ units of flow in $B$. Subsequently, we will show that the probability of this event is at least $0.95$ and this will prove the claim.

Since $\|\mathbf{s}\|_0 \leq \ell \leq \sqrt{d}$, we have $\Pr[\mathcal{E}_1] \geq 1 - 1/\sqrt{d}$. Since $\mathbf{x}$ is $(\ell, \log^{-5} d)$-up-persistent, $\Pr[\mathcal{E}_2] \geq 1 - \log^{-5} d$. Note that conditioned in $\mathcal{E}_1$, the distribution on $\mathbf{y}+\mathbf{s}$ is identical to $\mathcal{U}_\ell(\mathbf{y})$. Thus, since $\mathbf{y}$ is $(\ell, \log^{-5} d)$-up-persistent$\Pr[\mathcal{E}_3 \mid \mathcal{E}_1] \geq 1 - \log^{-5} d$. By a union bound

$$\Pr[\mathcal{E}_1 \wedge \mathcal{E}_2 \wedge \mathcal{E}_3] \geq 1 - 3\log^{-5} d. \tag{8}$$

To deal with $\mathcal{E}_4$, we bring in the non-redness of our edge $(\mathbf{x}, \mathbf{y})$. By definition,

$$\Pr_{\mathbf{z} \in I(\mathbf{x},\mathbf{y})} \Pr_{\mathbf{z}' \sim \mathcal{U}_\ell(\mathbf{z})}[\mathbf{z}' \text{ is not } \ell\text{-mzb}] \geq 0.99$$

In terms of shifts, we can express this bound as

$$\Pr_{\mathbf{z} \in I(\mathbf{x},\mathbf{y})} \Pr_{\mathbf{s} \sim \mathcal{US}_\ell(\mathbf{z})}[\mathbf{z}+\mathbf{s} \text{ is not } \ell\text{-mzb}] \geq 0.99$$

Since the probability of $\mathcal{E}_1$ is at least $1 - o(1)$, we have

$$\Pr_{\mathbf{z} \in I(\mathbf{x},\mathbf{y})} \Pr_{\mathbf{s} \sim \mathcal{US}_\ell(\mathbf{z})} [\mathbf{z} + \mathbf{s} \text{ is not } \ell\text{-mzb} \mid \mathcal{E}_1] \geq 0.98$$

Note that conditioned in $\mathcal{E}_1$, the distributions $\mathcal{US}_\ell(\mathbf{z})$ and $\mathcal{US}_\ell(\mathbf{x})$ are identical. Hence,

$$\Pr_{\mathbf{s} \sim \mathcal{US}_\ell(\mathbf{x})} \Pr_{\mathbf{z} \in I(\mathbf{x},\mathbf{y})} [\mathbf{z} + \mathbf{s} \text{ is not } \ell\text{-mzb} \mid \mathcal{E}_1] \geq 0.98$$

Let $X_\mathbf{s}$ be the fraction of points in $I(\mathbf{x} + \mathbf{s}, \mathbf{y} + \mathbf{s})$ that are not $\ell$-mzb. By linearity of expectation, $\mathbb{E}_\mathbf{s}[X_\mathbf{s} \mid \mathcal{E}_1] \geq 0.98$. Hence $\mathbb{E}_\mathbf{s}[1 - X_\mathbf{s} \mid \mathcal{E}_1] \leq 0.02$ and by Markov's inequality, $\Pr_\mathbf{s}[1 - X_\mathbf{s} > 0.5 \mid \mathcal{E}_1] \leq 1/50$. Hence, $\Pr_\mathbf{s}[X_\mathbf{s} \geq 0.5 \mid \mathcal{E}_1] \geq 49/50 = .98$. Since $\Pr[\mathcal{E}_1] = 1 - o(1)$, we have $\Pr[\mathcal{E}_4] = \Pr_\mathbf{s}[X_\mathbf{s} \geq 0.5] \geq 0.97$.

Combining with (8), we have $\Pr[\wedge_{j=1}^4 \mathcal{E}_j] \geq 0.96$. When $\wedge_{j=1}^4 \mathcal{E}_j$ occurs, the edge $(\mathbf{x}', \mathbf{y}')$ is a violated edge and at least half of $I(\mathbf{x}', \mathbf{y}')$ is not $\ell$-mzb. For $\mathbf{z}' \in I(\mathbf{x}', \mathbf{y}')$ that is not $\ell$-mzb, by definition $\Pr_{\mathbf{w} \sim \mathcal{D}_\ell(\mathbf{z}')}[f(\mathbf{w}) = 1] \geq 0.1$. Hence,

$$\Pr_{\mathbf{z}' \in_R I(\mathbf{x}',\mathbf{y}')} \Pr_{\mathbf{w} \sim \mathcal{D}_\ell(\mathbf{z}')} [f(\mathbf{w}) = 1] \geq 0.5 \times 0.1 \geq 0.01$$

We conclude that $(\mathbf{x}', \mathbf{y}')$ is blue, whenever $\wedge_{j=1}^4 \mathcal{E}_j$ occurs.

Stepping back, with probability at least 0.96 over the shift $\mathbf{s} \sim \mathcal{US}_\ell(\mathbf{x})$, the edge $(\mathbf{x} + \mathbf{s}, \mathbf{y} + \mathbf{s})$ is blue. Finally, since all points in $\boldsymbol{X}$ are 99-typical, we have $\Pr[\mathbf{x} \in \boldsymbol{H}_{99}] \geq 1 - (\varepsilon/d)^5$, and conditioned on this event we have $\mathbf{x}' \in \boldsymbol{H}_{100}$ since $\ell \ll \sqrt{d}$. Together, we get $\Pr[\mathcal{E}_5] \geq 1 - 2(\varepsilon/d)^5 \geq 0.99$. Thus, by a union bound $\Pr[\wedge_{j=1}^5 \mathcal{E}_j] \geq 0.95$ and so the amount of flow that $(\mathbf{x}, \mathbf{y})$ inserts is at least 0.95. □

Let $E' \subseteq B$ denote the set of blue edges which receive non-zero flow. Let $H(\boldsymbol{L}, \boldsymbol{R}, E')$ denote the bipartite graph on these edges. Since $\ell \leq \sqrt{d}/\log^5(d/\varepsilon)$, by the reversibility Lemma 3.24, $p_{\mathbf{x},\ell}(\mathbf{x}') \leq 2p_{\mathbf{x}',\ell}(\mathbf{x})$ for any $\mathbf{x} \in \boldsymbol{X}, \mathbf{x}' \in \boldsymbol{L}$ and $p_{\mathbf{y},\ell}(\mathbf{y}') \leq 2p_{\mathbf{y}',\ell}(\mathbf{y})$ for any $\mathbf{y} \in \boldsymbol{Y}, \mathbf{y}' \in \boldsymbol{R}$. Using this bound we're able to establish the desired capacity constraints on the flow as follows.

**Claim 7.2** (Edge Congestion)**.** *The total flow on an edge $(\mathbf{x}', \mathbf{y}') \in B$ is at most 2.*

*Proof.* By construction, the total flow on an edge $(\mathbf{x}', \mathbf{y}')$ is at most

$$\sum_{\mathbf{x} \in \boldsymbol{X}} p_{\mathbf{x},\ell}(\mathbf{x}') \leq 2 \sum_{\mathbf{x} \in \boldsymbol{X}} p_{\mathbf{x}',\ell}(\mathbf{x}) \leq 2$$

since $\sum_{\mathbf{x} \in \boldsymbol{X}} p_{\mathbf{x}',\ell}(\mathbf{x}) \leq 1$. □

**Claim 7.3** (Vertex Congestion)**.** *The following hold.*

1. *The total amount of flow through a vertex $\mathbf{x}' \in \boldsymbol{L}$ is at most $2D(\boldsymbol{X})$.*
2. *The total amount of flow through a vertex $\mathbf{y}' \in \boldsymbol{R}$ is at most $2D(\boldsymbol{Y})$.*
3. *For all $i \in [d]$, the total amount of $i$-flow through a vertex $\mathbf{x}' \in \boldsymbol{L}$ is at most $2\Gamma_i(\boldsymbol{X})$.*
4. *For all $i \in [d]$, the total amount of $i$-flow through a vertex $\mathbf{y}' \in \boldsymbol{R}$ is at most $2\Gamma_i(\boldsymbol{Y})$.*

*Proof.* The total flow through a vertex $\mathbf{x}' \in \boldsymbol{L}$ is at most

$$\sum_{(\mathbf{x},\mathbf{y})\in E} p_{\mathbf{x},\ell}(\mathbf{x}') \le D(\boldsymbol{X}) \sum_{\mathbf{x}\in \boldsymbol{X}} p_{\mathbf{x},\ell}(\mathbf{x}') \quad (\text{max degree of } \mathbf{x} \in \boldsymbol{X} \text{ is } D(\boldsymbol{X}))$$
$$\le 2D(\boldsymbol{X}) \sum_{\mathbf{x}\in[n]^d} p_{\mathbf{x}',\ell}(\mathbf{x}) \quad (\text{since } p_{\mathbf{x},\ell}(\mathbf{x}') \le 2p_{\mathbf{x}',\ell}(\mathbf{x}))$$
$$\le 2D(\boldsymbol{X}) \quad (\text{since } \sum_{\mathbf{x}\in[n]^d} p_{\mathbf{x}',\ell}(\mathbf{x}) \le 1)$$

and an analogous argument proves (2). For a coordinate $i \in [d]$, let $E_i \subseteq E$ denote the set of $i$-edges in $G$. The total $i$-flow through a vertex $\mathbf{x}' \in \boldsymbol{L}$ is at most

$$\sum_{(\mathbf{x},\mathbf{y})\in E_i} p_{\mathbf{x}}(\mathbf{x}') \le \Gamma_i(\boldsymbol{X}) \sum_{\mathbf{x}\in \boldsymbol{X}} p_{\mathbf{x},\ell}(\mathbf{x}') \quad (\text{max } i\text{-degree of } \mathbf{x} \in \boldsymbol{X} \text{ is } \Gamma_i(\boldsymbol{X}))$$
$$\le 2\Gamma_i(\boldsymbol{X}) \sum_{\mathbf{x}\in[n]^d} p_{\mathbf{x}',\ell}(\mathbf{x}) \quad (\text{since } p_{\mathbf{x},\ell}(\mathbf{x}') \le 2p_{\mathbf{x}',\ell}(\mathbf{x}))$$
$$\le 2\Gamma_i(\boldsymbol{X}) \quad (\text{since } \sum_{\mathbf{x}\in[n]^d} p_{\mathbf{x}',\ell}(\mathbf{x}) \le 1)$$

and an analogous argument proves (4). □

By Claim 7.1 and the fact that at least half the edges in $G$ are not red, the total amount of flow is at least $m(G)/3$ and this flow satisfies the constraints listed in Claim 7.2 and Claim 7.3. Thus, dividing by 2 yields a flow of value $m(G)/6$ satisfying the following.

1. The flow on every edge is at most 1.
2. The total flow through any vertex in $\boldsymbol{L}$ is at most $D(\boldsymbol{X})$. The total $i$-flow through any vertex in $\boldsymbol{L}$ is at most $\Gamma_i(\boldsymbol{X})$.
3. The total flow through any vertex in $\boldsymbol{R}$ is at most $D(\boldsymbol{Y})$. The total $i$-flow through any vertex in $\boldsymbol{R}$ is at most $\Gamma_i(\boldsymbol{Y})$.

By integrality of flow, there exists an integral flow of at least $\lfloor m(G)/6 \rfloor \ge m(G)/7$ units satisfying the same capacity constraints. By item (1) above, the integral flow is a subgraph containing at least $m/7$ edges and satisfying the desired constraints listed in the lemma statement. □

# 8 Proof of the 'Persist-or-Blow-Up' Lemma, Lemma 6.7

Let us recall the 'Persist-or-Blow-Up' lemma.

**Lemma 6.7** (Persist-or-Blow-up Lemma). *Consider a violation subgraph $G = (\boldsymbol{X}, \boldsymbol{Y}, E)$ such that all vertices in $G$ are $c$-typical where $c \le 99$ and $(\ell, \log^{-5} d)$-up persistent where $1 \le \ell \le \sqrt{d}/\log^5(d/\varepsilon)$. Then, there exists a violation subgraph $G' = (\boldsymbol{X}', \boldsymbol{Y}', E')$ where all vertices are $(c + \frac{\ell}{\sqrt{d}})$-typical and satisfying one of the following conditions.*

1. *Down-persistent case:*

(a) *All vertices in* $\boldsymbol{X}'$ *are* $(\ell, 0.6)$*-down persistent.*

(b) $m(G') \geq m(G)/\log^5 d$.

(c) $D(\boldsymbol{X}') \leq D(\boldsymbol{X})$, *and* $\forall i \in [d], \Gamma_i(\boldsymbol{X}') \leq \Gamma_i(\boldsymbol{X})$

(d) $D(\boldsymbol{Y}') \leq D(\boldsymbol{Y})$, *and* $\forall i \in [d]$, $\Gamma_i(\boldsymbol{Y}') \leq \Gamma_i(\boldsymbol{Y})$.

2. *Blow-up case:*

(a) $m(G') \geq 2(1 - 3\log^{-3} d) \cdot m(G)$.

(b) $D(\boldsymbol{X}') \leq D(\boldsymbol{X})$, *and* $\forall i \in [d], \Gamma_i(\boldsymbol{X}') \leq \Gamma_i(\boldsymbol{X})$

(c) $D(\boldsymbol{Y}') \leq 2D(\boldsymbol{Y})$, *and* $\forall i \in [d]$, $\Gamma_i(\boldsymbol{Y}') \leq 2\Gamma_i(\boldsymbol{Y})$.

We first recall the definition of $p_{\mathbf{x},\ell}(\mathbf{x}')$ in Definition 3.23. For a fixed $\mathbf{x}$, consider the process of sampling a hypercube $\boldsymbol{H} \sim \mathbb{H}(\mathbf{x})$ and then sampling $\mathbf{z} \sim \mathcal{U}_{\boldsymbol{H},\ell}(\mathbf{x})$. Recall from Fact 3.11 that this is one of three equivalent ways of expressing our random walk distribution. Given $\mathbf{x}, \mathbf{x}', \ell$, we have

$$p_{\mathbf{x},\ell}(\mathbf{x}') = \Pr\left[\mathbf{x}, \mathbf{x}' \in \boldsymbol{H}_{100} \text{ and } \mathbf{z} = \mathbf{x}'\right].$$

We use these values to set up a flow problem as follows. For every edge $(\mathbf{x}, \mathbf{y})$ of $G$ and $\mathbf{s} \in \text{supp}(\mathcal{US}_\ell(\mathbf{x}))$, if $e = (\mathbf{x}+\mathbf{s}, \mathbf{y}+\mathbf{s})$ is a violation, then we put $p_{\mathbf{x}}(\mathbf{x}+\mathbf{s})$ units of flow on $e$. The flow, denoted $\mathcal{F}$, is supported on a violation subgraph $G' = (\boldsymbol{X}', \boldsymbol{Y}', E)$. Note that by Claim 3.22, all vertices in $G'$ are $(c + \frac{\ell}{\sqrt{d}})$-typical.

**Claim 8.1.** *Every edge of* $G$ *inserts at least* $1 - \log^{-4} d$ *units of flow.*

*Proof.* The proof of this claim is similar to that of Claim 7.1. Fix an edge $(\mathbf{x}, \mathbf{y}) \in G$ and let this be an $i$-edge. Generate $\boldsymbol{H} \sim \mathbb{H}(\mathbf{x})$ and a shift $\mathbf{s} \sim \mathcal{US}_{\boldsymbol{H},\ell}(\mathbf{x})$, and let $\mathbf{x}' = \mathbf{x} + \mathbf{s}$ and $\mathbf{y}' = \mathbf{y} + \mathbf{x}$. Consider the events: (i) $\mathcal{E}_1$: $\mathbf{s}_i = 0$, (ii) $\mathcal{E}_2$: $f(\mathbf{x}') = 1$, (iii) $\mathcal{E}_3$: $f(\mathbf{y}') = 0$, (iv)$\mathcal{E}_4$: $\mathbf{x}, \mathbf{x}' \in \boldsymbol{H}_{100}$. Note that the total flow inserted by $(\mathbf{x}, \mathbf{y})$ is at least $\Pr[\wedge_{i=1}^4 \mathcal{E}_i]$. $\Pr[\mathcal{E}_1] \geq 1 - 1/\sqrt{d}$, since $\|\mathbf{s}\|_0 \leq \ell \leq \sqrt{d}$. Since $\mathbf{x}, \mathbf{y}$ are both $(\ell, \log^{-5} d)$-up-persistent and $f(\mathbf{x}) = 1$, $f(\mathbf{y}) = 0$, we get $\Pr[\mathcal{E}_2], \Pr[\mathcal{E}_3] \geq 1 - \frac{1}{\log^5 d}$. Finally, since $\mathbf{x}$ is 99-typical, with probability $1 - (\varepsilon/d)^5$ we have $\mathbf{x} \in \boldsymbol{H}_{99}$ which implies $\mathbf{x}' \in \boldsymbol{H}_{100}$ since $\ell \ll \sqrt{d}$. Thus by a union bound, $\Pr[\wedge_{i=1}^5 \mathcal{E}_i] \geq 1 - 2\log^{-5} d - 1/\sqrt{d} - (\varepsilon/d)^5 \geq 1 - \log^{-4} d$. $\square$

**Claim 8.2** (Edge Congestion)**.** *The flow on any edge* $(\mathbf{x}', \mathbf{y}')$ *is at most* $\sum_{\mathbf{x} \in \boldsymbol{X}} p_{\mathbf{x},\ell}(\mathbf{x}') \leq (1 + \log^{-3} d)$.

*Proof.* Consider an edge $(\mathbf{x}', \mathbf{y}')$, which receives flow from some $(\mathbf{x}, \mathbf{y})$ in $G$. Flow is inserted by translations of edges, so $\mathbf{y} - \mathbf{x} = \mathbf{y}' - \mathbf{x}'$. Hence, for a given $\mathbf{x}$, there exists a unique $\mathbf{y}$ such that $(\mathbf{x}, \mathbf{y})$ inserts flow on $(\mathbf{x}', \mathbf{y}')$. By construction, the flow inserted is $p_{\mathbf{x},\ell}(\mathbf{x}')$. Thus, the total flow that $(\mathbf{x}', \mathbf{y}')$ receives is at most $\sum_{\mathbf{x} \in \boldsymbol{X}} p_{\mathbf{x},\tau}(\mathbf{x}')$. The RHS bound holds by Lemma 3.24 and observing that $\sum_{\mathbf{x} \in \boldsymbol{X}} p_{\mathbf{x}',\ell}(\mathbf{x}) \leq 1$. $\square$

**Claim 8.3** (Vertex Congestion)**.** *The following hold.*

1. *For any* $\mathbf{x}' \in \boldsymbol{X}'$*, the total flow on edges incident to* $\mathbf{x}'$ *is at most*

$$D(\boldsymbol{X}) \sum_{\mathbf{x} \in \boldsymbol{X}} p_{\mathbf{x},\ell}(\mathbf{x}') \leq D(\boldsymbol{X})(1 + \log^{-3} d).$$

2. *For any* $\mathbf{x}' \in \boldsymbol{X}'$*, the total* $i$*-flow on edges incident to* $\mathbf{x}'$ *is at most*

$$\Gamma_i(\boldsymbol{X}) \sum_{\mathbf{x} \in \boldsymbol{X}} p_{\mathbf{x},\ell}(\mathbf{x}') \leq \Gamma_i(\boldsymbol{X})(1 + \log^{-3} d).$$

3. *For any $\mathbf{y}' \in \boldsymbol{Y}'$, the total flow on edges incident to $\mathbf{y}'$ is at most*

$$D(\boldsymbol{Y}) \sum_{\mathbf{y} \in \boldsymbol{Y}} p_{\mathbf{y},\ell}(\mathbf{y}') \leq D(\boldsymbol{Y})(1 + \log^{-3} d).$$

4. *For any $\mathbf{y}' \in \boldsymbol{Y}'$, the total $i$-flow on edges incident to $\mathbf{y}'$ is at most*

$$\Gamma_i(\boldsymbol{Y}) \sum_{\mathbf{y} \in \boldsymbol{Y}} p_{\mathbf{y},\ell}(\mathbf{y}') \leq \Gamma_i(\boldsymbol{Y})(1 + \log^{-3} d).$$

*Proof.* Consider $\mathbf{x}' \in \boldsymbol{X}'$. All the $i$-flow inserted on edges incident to $\mathbf{x}'$ comes from $i$-edges $(\mathbf{x}, \mathbf{y})$ in $G$. Every $i$-edge in $G$ inserts flow on at most a single edge incident to $\mathbf{x}'$ and there are at most $\Gamma_i(\boldsymbol{X})$ $i$-edges incident to any vertex $\mathbf{x} \in \boldsymbol{X}$. Hence, the total $i$-flow inserted by a $\mathbf{x} \in \boldsymbol{X}$ through $\mathbf{x}'$ is at most $\Gamma_i(\boldsymbol{X}) \cdot p_{\mathbf{x},\tau}(\mathbf{x}')$. Thus, summing over all $\mathbf{x} \in \boldsymbol{X}$ and using the reversibility Lemma 3.24 shows that the total $i$-flow on edges incident to $\mathbf{x}'$ is at most

$$\Gamma_i(\boldsymbol{X}) \sum_{\mathbf{x} \in \boldsymbol{X}} p_{\mathbf{x},\tau}(\mathbf{x}') \leq (1 + \log^{-3} d)\Gamma_i(\boldsymbol{X}) \sum_{\mathbf{x} \in \boldsymbol{X}} p_{\mathbf{x}',\tau}(\mathbf{x}) \leq (1 + \log^{-3} d)\Gamma_i(\boldsymbol{X})$$

and this proves (2). The proof of (1) is identical, with $D(\boldsymbol{X})$ replacing $\Gamma_i(\boldsymbol{X})$, and statements (3) and (4) have analogous proofs. $\square$

We now come to a key definition in our analysis.

**Definition 8.4** (Heavy Vertices)**.** *A vertex $\mathbf{x}' \in \boldsymbol{X}'$ is called* heavy *if it satisfies any of the following.*

1. *There is an edge $(\mathbf{x}', \mathbf{y}')$ receiving at least $1/2$ units of flow.*
2. *The total flow on edges incident to $\mathbf{x}'$ is at least $D(\boldsymbol{X})/2$.*
3. *There exists $i \in [d]$ such that the total $i$-flow on edges incident to $\mathbf{x}'$ is at least $\Gamma_i(\boldsymbol{X})/2$.*

We refer to the flow passing through heavy vertices as the *heavy flow*.

**Claim 8.5.** *All heavy vertices are $(\ell, 0.6)$-down persistent.*

*Proof.* Consider a heavy vertex $\mathbf{x}'$. That is, $\mathbf{x}'$ satisfies one of the three conditions listed in Definition 8.4. Suppose it satisfies the first condition: there is some violated edge $(\mathbf{x}', \mathbf{y}')$ receiving at least $1/2$ units of flow. By Claim 8.2, the total flow on $(\mathbf{x}', \mathbf{y}')$ is at most $\sum_{\mathbf{x} \in \boldsymbol{X}} p_{\mathbf{x},\ell}(\mathbf{x}')$. Hence, $\sum_{\mathbf{x} \in \boldsymbol{X}} p_{\mathbf{x},\ell}(\mathbf{x}') \geq 1/2$. In fact, observe that we can prove the exact same inequality if $\mathbf{x}'$ satisfies the second or third condition of Definition 8.4, by using the upper bound given by the LHS of items (1) and (2), respectively, of Claim 8.3. Now, applying the reversibility Lemma 3.24, we have $(1 + \log^{-3} d) \sum_{\mathbf{x} \in \boldsymbol{X}} p_{\mathbf{x}',\ell}(\mathbf{x}) \geq 1/2$. Note that $f(\mathbf{x}) = 1$ for all $\mathbf{x} \in \boldsymbol{X}$. Hence,

$$\Pr_{\mathbf{z} \sim \mathcal{D}_\ell(\mathbf{x}')}[f(\mathbf{z}) = 1] \geq \sum_{\mathbf{x} \in \boldsymbol{X}} p_{\mathbf{x}',\ell}(\mathbf{x}) \geq \frac{1}{2(1 + \log^{-3} d)} \geq 0.4 \tag{9}$$

and so $\mathbf{x}'$ is $(\ell, 0.6)$-down-persistent. $\square$

We are now set up to complete the proof. For convenience, we use $m$ to denote $m(G)$. We refer to the flow on edges incident to heavy vertices as the *heavy flow*. We let $G_H(\boldsymbol{X}_H, \boldsymbol{Y}_H, E_H)$ denote the bipartite graph of all edges incident to heavy vertices, that is, $\boldsymbol{X}_H$ is the set of all heavy vertices. We refer to the flow on edges incident to non-heavy vertices as the *light flow*. We let $G_L(\boldsymbol{X}_L, \boldsymbol{Y}_L, E_L)$ denote the bipartite graph of all edges incident to non-heavy vertices, that is, $\boldsymbol{X}_L = \boldsymbol{X}' \setminus \boldsymbol{X}_H$ is the set of all non-heavy vertices. We split into two cases based on the amount of heavy flow.

### 8.1 Case 1: The total amount of heavy flow is at least $\frac{m}{\log^4 d}$

Note that by Claim 8.5, all vertices in $\boldsymbol{X}_H$ are $(\ell, 0.6)$-down persistent.

By Claim 8.2 and Claim 8.3, the heavy flow satisfies the following capacity constraints.

1. The flow on every edge is at most $(1 + \log^{-3} d)$.
2. For every $\mathbf{x}' \in \boldsymbol{X}_H$, the total flow on edges incident to $\mathbf{x}'$ is at most $D(\boldsymbol{X})(1 + \log^{-3} d)$ and the total $i$-flow on edges incident to $\mathbf{x}'$ is at most $\Gamma_i(\boldsymbol{X})(1 + \log^{-3} d)$.
3. For every $\mathbf{y}' \in \boldsymbol{Y}_H$, the total flow on edges incident to $\mathbf{y}'$ is at most $D(\boldsymbol{Y})(1 + \log^{-3} d)$ and the total $i$-flow on edges incident to $\mathbf{y}'$ is at most $\Gamma_i(\boldsymbol{Y})(1 + \log^{-3} d)$.

Let us divide the flow by $(1 + \log^{-3} d)$. Thus, we now have at least $\frac{m}{(1+\log^{-3} d)\log^4 d} \geq \frac{m}{\log^5 d}$ units of flow satisfying the following capacity constraints.

1. The flow on every edge is at most one.
2. For every $\mathbf{x}' \in \boldsymbol{X}_H$, the total flow on edges incident to $\mathbf{x}'$ is at most $D(\boldsymbol{X})$ and the total $i$-flow on edges incident to $\mathbf{x}'$ is at most $\Gamma_i(\boldsymbol{X})$.
3. For every $\mathbf{y}' \in \boldsymbol{Y}_H$, the total flow on edges incident to $\mathbf{y}'$ is at most $D(\boldsymbol{Y})$ and the total $i$-flow on edges incident to $\mathbf{y}'$ is at most $\Gamma_i(\boldsymbol{Y})$.

By integrality of flow, there is an integral flow of at least $\frac{m}{\log^5 d}$ units satisfying the above constraints. By condition (1) above, this integral flow is a subgraph of $G_H$ with at least $\frac{m}{\log^5 d}$ edges, and satisfying the degree bounds listed in (1c) and (1d) of the lemma statement. Thus, this subgraph satisfies case (1) of the lemma statement.

### 8.2 Case 2: The total amount of heavy flow is at most $\frac{m}{\log^4 d}$

By Claim 8.1, the total flow is at least $m(1 - \log^{-4} d)$ units. Thus, after removing the heavy flow, the remaining light flow is at least $m(1 - 2\log^{-4} d)$ units. The light flow satisfies the following capacity constraints.

1. Every edge has at most $1/2$ units of flow.
2. For every $\mathbf{x}' \in \boldsymbol{X}_L$, the total flow on edges incident to $\mathbf{x}'$ is at most $D(\boldsymbol{X})/2$ and the total $i$-flow on edges incident to $\mathbf{x}'$ is at most $\Gamma_i(\boldsymbol{X})/2$.
3. For every $\mathbf{y}' \in \boldsymbol{Y}_L$, the total flow on edges incident to $\mathbf{y}'$ is at most $(1 + \log^{-3} d)D(\boldsymbol{Y})$ and the total $i$-flow on edges incident to $\mathbf{y}'$ is at most $(1 + \log^{-3} d)\Gamma_i(\boldsymbol{Y})$.

Items (1) and (2) are simply by Definition 8.4 since all vertices in $\boldsymbol{X}_L$ are not heavy. Item (3) follows from RHS bound on the vertex congestion in Claim 8.3.

We now by rescale the flow by multiplying it by $\frac{2}{1+\log^{-3} d}$. We now have $2m\frac{(1-2\log^{-4} d)}{1+\log^{-3} d} \geq 2m(1 - 2\log^{-3} d)$ units of flow with the following capacity constraints:

1. Every edge has at most one unit of flow.

2. For every $\mathbf{x}' \in \boldsymbol{X}_L$, the total flow on edges incident to $\mathbf{x}'$ is at most $D(\boldsymbol{X})$ and the total $i$-flow on edges incident to $\mathbf{x}'$ is at most $\Gamma_i(\boldsymbol{X})$.

3. For every $\mathbf{y}' \in \boldsymbol{Y}_L$, the total flow on edges incident to $\mathbf{y}'$ is at most $2D(\boldsymbol{Y})$ and the total $i$-flow on edges incident to $\mathbf{y}'$ is at most $2\Gamma_i(\boldsymbol{Y})$.

By integrality of flow, we obtain an integral flow of at least $\lfloor 2m(1 - 3\log^{-4} d) \rfloor \geq 2m(1 - 3\log^{-3} d)$ units satisfying the same constraints listed above. In particular, the flow on any edge is at most one and so the integral flow is a violation subgraph with at least $2m(1 - 3\log^{-3} d)$ edges and satisfying the degree bounds listed in case (2) of the lemma statement.

## References


[AC06] Nir Ailon and Bernard Chazelle. Information theory in property testing and monotonicity testing in higher dimension. *Information and Computation*, 204(11):1704–1717, 2006. 7

[ACCL07] Nir Ailon, Bernard Chazelle, Seshadhri Comandur, and Ding Liu. Estimating the distance to a monotone function. *Random Structures Algorithms*, 31(3):371–383, 2007. Prelim. version in Proc., RANDOM 2004. 7

[BB21] Aleksandrs Belovs and Eric Blais. A polynomial lower bound for testing monotonicity. *SIAM Journal on Computing (SICOMP)*, 50(3):406–433, 2021. Prelim. version in Proc., STOC 2016. 7

[BBM12] Eric Blais, Joshua Brody, and Kevin Matulef. Property testing lower bounds via communication complexity. *Computational Complexity*, 21(2):311–358, 2012. Prelim. version in Proc., CCC 2011. 7

[BCS18] Hadley Black, Deeparnab Chakrabarty, and C. Seshadhri. A $o(d) \cdot \text{polylog}(n)$ monotonicity tester for Boolean functions over the hypergrid $[n]^d$. In *Proceedings, ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2018. 1, 7, 8

[BCS20] Hadley Black, Deeparnab Chakrabarty, and C. Seshadhri. Domain reduction: A $o(d)$ tester for boolean functions in $d$-dimensions. In *Proceedings, ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2020. 1, 7, 8, 10

[BCS23] Hadley Black, Deeparnab Chakrabarty, and C. Seshadhri. Directed isoperimetric theorems for boolean functions on the hypergrid and an $\widetilde{O}(n\sqrt{d})$ monotonicity tester. In *Proceedings, ACM Symposium on Theory of Computing (STOC)*, 2023. 1, 2, 3, 6, 7, 8, 11, 12, 21

[BCSM12] Jop Briët, Sourav Chakraborty, David García Soriano, and Ari Matsliah. Monotonicity testing and shortest-path routing on the cube. *Combinatorica*, 32(1):35–53, 2012. 7

[BGJ+12] Arnab Bhattacharyya, Elena Grigorescu, Madhav Jha, Kyoming Jung, Sofya Raskhodnikova, and David Woodruff. Lower bounds for local monotonicity reconstruction from transitive-closure spanners. *SIAM Journal on Discrete Mathematics (SIDMA)*, 26(2):618–646, 2012. Prelim. version in Proc., RANDOM 2010. 7

[Bha08] Arnab Bhattacharyya. A note on the distance to monotonicity of boolean functions. Technical Report 012, Electronic Colloquium on Computational Complexity (ECCC), 2008. 7

[BKKM23] Mark Braverman, Subhash Khot, Guy Kindler, and Dor Minzer. Improved monotonicity testers via hypercube embeddings. In *Innovations in Theoretical Computer Science (ITCS)*, pages 25:1–25:24, 2023. 1, 7, 8, 11

[BKR20] Hadley Black, Iden Kalemaj, and Sofya Raskhodnikova. Isoperimetric inequalities for real-valued functions with applications to monotonicity testing. *arXiv*, abs/2011.09441, 2020. 7

[BRY14a] Piotr Berman, Sofya Raskhodnikova, and Grigory Yaroslavtsev. $L_p$-testing. In *Proceedings, ACM Symposium on Theory of Computing (STOC)*, 2014. 7, 10

[BRY14b] Eric Blais, Sofya Raskhodnikova, and Grigory Yaroslavtsev. Lower bounds for testing properties of functions over hypergrid domains. In *Proceedings, IEEE Conference on Computational Complexity (CCC)*, 2014. 7

[CDJS17] Deeparnab Chakrabarty, Kashyap Dixit, Madhav Jha, and C Seshadhri. Property testing on product distributions: Optimal testers for bounded derivative properties. *ACM Trans. on Algorithms (TALG)*, 13(2):1–30, 2017. Prelim. version in Proc., SODA 2015. 7

[CDST15] Xi Chen, Anindya De, Rocco A. Servedio, and Li-Yang Tan. Boolean function monotonicity testing requires (almost) $O(n^{1/2})$ non-adaptive queries. In *Proceedings, ACM Symposium on Theory of Computing (STOC)*, 2015. 7

[CS13] Deeparnab Chakrabarty and C. Seshadhri. Optimal bounds for monotonicity and Lipschitz testing over hypercubes and hypergrids. In *Proceedings, ACM Symposium on Theory of Computing (STOC)*, 2013. 7

[CS14] Deeparnab Chakrabarty and C. Seshadhri. An $o(n)$ monotonicity tester for Boolean functions over the hypercube. *SIAM Journal on Computing (SICOMP)*, 45(2):461–472, 2014. Prelim. version in Proc., STOC 2013. 2, 7

[CST14] Xi Chen, Rocco A. Servedio, and Li-Yang. Tan. New algorithms and lower bounds for monotonicity testing. In *Proceedings, IEEE Symposium on Foundations of Computer Science (FOCS)*, 2014. 7

[CWX17] Xi Chen, Erik Waingarten, and Jinyu Xie. Beyond Talagrand: New lower bounds for testing monotonicity and unateness. In *Proceedings, ACM Symposium on Theory of Computing (STOC)*, 2017. 1, 7

[DGL+99] Yevgeny Dodis, Oded Goldreich, Eric Lehman, Sofya Raskhodnikova, Dana Ron, and Alex Samorodnitsky. Improved testing algorithms for monotonicity. *Proceedings, International Workshop on Randomization and Computation (RANDOM)*, 1999. 7, 10

[EKK+00] Funda Ergun, Sampath Kannan, Ravi Kumar, Ronitt Rubinfeld, and Mahesh Viswanathan. Spot-checkers. *J. Comput. System Sci.*, 60(3):717–751, 2000. Prelim. version in Proc., STOC 1998. 7

[Fis04] Eldar Fischer. On the strength of comparisons in property testing. *Information and Computation*, 189(1):107–116, 2004. 7

[FLN+02] Eldar Fischer, Eric Lehman, Ilan Newman, Sofya Raskhodnikova, and Ronitt Rubinfeld. Monotonicity testing over general poset domains. *Proceedings, ACM Symposium on Theory of Computing (STOC)*, 2002. 7

[FR10] Shahar Fattal and Dana Ron. Approximating the distance to monotonicity in high dimensions. *ACM Trans. on Algorithms (TALG)*, 6(3), 2010. 7

[GGL+00] Oded Goldreich, Shafi Goldwasser, Eric Lehman, Dana Ron, and Alex Samordinsky. Testing monotonicity. *Combinatorica*, 20:301–337, 2000. Prelim. version in Proc., FOCS 1998. 7

[HK03] Shirley Halevy and Eyal Kushilevitz. Distribution-free property testing. *Proceedings, International Workshop on Randomization and Computation (RANDOM)*, 2003. 7

[HK08] Shirley Halevy and Eyal Kushilevitz. Testing monotonicity over graph products. *Random Structures Algorithms*, 33(1):44–67, 2008. Prelim. version in Proc., ICALP 2004. 7

[HY22] Nathaniel Harms and Yuichi Yoshida. Downsampling for testing and learning in product distributions. In *Proceedings, International Colloquium on Automata, Languages and Programming (ICALP)*, volume 229, pages 71:1–71:19, 2022. 1, 7, 10

[KMS18] Subhash Khot, Dor Minzer, and Muli Safra. On monotonicity testing and boolean isoperimetric-type theorems. *SIAM Journal on Computing*, 47(6):2238–2276, 2018. Prelim. version in Proc., FOCS 2015. 1, 2, 3, 6, 7, 12, 14, 21, 40, 41

[LR01] Eric Lehman and Dana Ron. On disjoint chains of subsets. *Journal of Combinatorial Theory, Series A*, 94(2):399–404, 2001. 7

[Ras99] Sofya Raskhodnikova. Monotonicity testing. *Masters Thesis, MIT*, 1999. 7

[RRS+12] Dana Ron, Ronitt Rubinfeld, Muli Safra, Alex Samorodnitsky, and Omri Weinstein. Approximating the influence of monotone boolean functions in $O(\sqrt{n})$ query complexity. *ACM Trans. Comput. Theory*, 4(4):11:1–11:12, 2012. Prelim. version in Proc., RANDOM 2011. 7

[SS08] Michael E. Saks and C. Seshadhri. Parallel monotonicity reconstruction. In *Proceedings, ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2008. 7

# A Deferred Proofs

## A.1 Equivalence of the Walk Distributions: Proof of Fact 3.11

*Proof.* Fix a pair $(u, v)$ in $[n]^d$ where $u \preceq v$. We will show that the probability of sampling this pair from each distribution is the same. Let $S = \{i \in [d] \colon v_i > u_i\}$. Note that $u_j = v_j$ for all $j \neq S$. The probability of sampling the pair $(u, v)$ from the distribution described in item (1) of Fact 3.11 is computed as follows.

$$\Pr_{\mathbf{x}\in_R[n]^d,\, \mathbf{y}\sim\mathcal{U}_\tau(\mathbf{x})}[(\mathbf{x},\mathbf{y}) = (u,v)] \quad = \frac{1}{n^d} \sum_{R\supseteq S\,:\,|R|=\tau} \binom{d}{\tau}^{-1} \prod_{i\in S} \Pr[c_i = v_i \mid \mathbf{x} = u] \prod_{i\in R\setminus S} \Pr[c_i \leq u_i \mid \mathbf{x} = u]. \tag{10}$$

Recall the distribution of $q_i, I_i, c_i$ from Definition 2.1. Consider $i \in S$ and let $d_i := \min(v_i - u_i, n - (v_i - u_i))$. Note that conditioned on $q_i$, the total number of intervals $I_i \ni u_i$ is $2^{q_i}$ and the number of such intervals that contain $v_i$ is $\max(0, 2^{q_i} - d_i)$. Thus, we have

$$i \in S \implies \Pr[c_i = v_i \mid \mathbf{x} = u] = \mathbb{E}_{q_i}\Big[\Pr_{I_i}[v_i \in I_i] \Pr_{c_i \in I_i}[c_i = v_i \mid v_i \in I_i]\Big]$$
$$= \frac{1}{\log n} \sum_{q\colon 2^{q_i} \geq d_i} \frac{2^{q_i} - d_i}{2^{q_i}} \cdot \frac{1}{2^{q_i} - 1} = \frac{1}{2} \cdot \mathbb{E}_{q_i}\left[\frac{\max(0, 2^{q_i} - d_i)}{\binom{2^{q_i}}{2}}\right]. \tag{11}$$

For an interval $I_i \ni u_i$, let $I_{i,u_i}$ denote the prefix of $I_i$ preceding (not including) $u_i$. Note that conditioned on an interval $I_i \ni u_i$, the probability of choosing $c_i \leq u_i$ is $|I_{i,u_i}|/(2^{q_i} - 1)$. Thus, we have

$$i \in R \setminus S \implies \Pr[c_i \leq u_i \mid \mathbf{x} = u] = \mathbb{E}_{q_i}\left[\frac{1}{2^{q_i} - 1} \cdot \mathbb{E}_{I_i \ni u_i}[|I_{i,u_i}|]\right] \tag{12}$$

We now compute the probability of sampling $(u, v)$ from the distribution described in item (2) of Fact 3.11. Recall the distribution of $q_i, I_i, a_i, b_i$ from Definition 3.8. For $i \in [d]$, let $\mathcal{E}_i$ be the event that $a_i = u_i$ or $b_i = u_i$. Note that

$$\Pr[\mathcal{E}_i] = \mathbb{E}_{q_i}\left[\Pr_{I_i}[I_i \ni u_i] \Pr_{a_i < b_i \in I_i}[u_i \in \{a_i, b_i\} \mid u_i \in I_i]\right] = \mathbb{E}_{q_i}\left[\frac{2^{q_i}}{n} \cdot \frac{2}{2^{q_i}}\right] = \frac{2}{n}$$

Let $\mathcal{E}_u$ denote the event that $\mathbf{x} = u$. We have

$$\Pr[\mathcal{E}_u] = \prod_{i=1}^{d} \Pr[\mathcal{E}_i] \cdot \frac{1}{2^d} = \left(\frac{2}{n}\right)^d \frac{1}{2^d} = \frac{1}{n^d}. \tag{13}$$

Let $\mathcal{E}_v$ denote the event that $\mathbf{y} = v$. We have

$$\Pr\left[\mathcal{E}_v \mid \mathcal{E}_u\right] = \sum_{R \supseteq S\,:\,|R| = \tau} \binom{d}{\tau}^{-1} \prod_{i \in S} \Pr[a_i = u_i \text{ and } b_i = v_i \mid \mathcal{E}_u] \cdot \prod_{i \in R \setminus S} \Pr[b_i = u_i \mid \mathcal{E}_u] \tag{14}$$

Fix an $i \in S$ and recall $d_i := \min(v_i - u_i, n - (v_i - u_i))$. We have

$$\Pr[a_i = u_i \text{ and } b_i = v_i \mid \mathcal{E}_u] = \Pr[a_i = u_i \text{ and } b_i = v_i \mid \mathcal{E}_i] = \frac{\Pr[a_i = u_i \text{ and } b_i = v_i]}{\Pr[\mathcal{E}_i]}$$

where the numerator is

$$\Pr[a_i = u_i \text{ and } b_i = v_i] = \mathbb{E}_{q_i}\left[\Pr_{I_i}\left[I_i \supseteq [u_i, v_i]\right] \cdot \binom{2^{q_i}}{2}^{-1}\right] = \mathbb{E}_{q_i}\left[\frac{\max(0, 2^{q_i} - d_i)}{n \cdot \binom{2^{q_i}}{2}}\right]$$

and so

$$i \in S \implies \Pr[a_i = u_i \text{ and } b_i = v_i \mid \mathcal{E}_u] = \frac{1}{2} \cdot \mathbb{E}_{q_i}\left[\frac{\max(0, 2^{q_i} - d_i)}{\binom{2^{q_i}}{2}}\right] \tag{15}$$

which is equal to the probability computed in (11).

Now fix an $i \in R \setminus S$. Recall the definition of $I_{i,u_i}$. We have

$$\Pr[b_i = u_i \mid \mathcal{E}_u] = \Pr[b_i = u_i \mid \mathcal{E}_i] = \frac{\Pr[b_i = u_i]}{\Pr[\mathcal{E}_i]}$$

where

$$\Pr[b_i = u_i] = \mathbb{E}_{q_i}\mathbb{E}_{I_i}\left[\mathbf{1}(u_i \in I_i)\frac{|I_{u_i}|}{\binom{2^{q_i}}{2}}\right] = \mathbb{E}_{q_i}\left[\frac{1}{n}\sum_{I_i \ni u_i}|I_{i,u_i}|\binom{2^{q_i}}{2}^{-1}\right] = \frac{2}{n}\mathbb{E}_{q_i}\left[\frac{1}{2^{q_i-1}}\cdot \mathbb{E}_{I_i \ni u_i}[|I_{i,u_i}|]\right]$$

and so recalling that $\Pr[\mathcal{E}_i] = 2/n$ we have

$$i \notin R \setminus S \implies \Pr[b_i = u_i \mid \mathcal{E}_u] = \mathbb{E}_{q_i}\left[\frac{1}{2^{q_i-1}}\cdot \mathbb{E}_{I_i \ni u_i}[|I_{i,u_i}|]\right] \tag{16}$$

which is equal to the probability computed in (12). Combining (10), (11), (12), (13), (14), (15), (16), we have

$$\Pr_{\boldsymbol{H}\sim\mathbb{H}}\Pr_{\mathbf{x}\in_R \boldsymbol{H},\, \mathbf{y}\sim\mathcal{U}_{\boldsymbol{H},\tau}(\mathbf{x})}[(\mathbf{x},\mathbf{y}) = (u,v)] = \Pr[\mathcal{E}_u]\cdot\Pr[\mathcal{E}_v \mid \mathcal{E}_u] = \Pr_{\mathbf{x}\in_R [n]^d,\, \mathbf{y}\sim\mathcal{U}_\tau(\mathbf{x})}[(\mathbf{x},\mathbf{y}) = (u,v)]$$

and this proves that (1) and (2) of Fact 3.11 are equivalent.

To show equivalence of (1) and (3), note that we only need to show that

$$\Pr_{\boldsymbol{H}\sim\mathbb{H}(u),\, \mathbf{y}\sim\mathcal{U}_{\boldsymbol{H},\tau}(u)}[\mathbf{y} = v] = \Pr_{\mathbf{y}\sim\mathcal{U}_\tau(u)}[\mathbf{y} = v] \tag{17}$$

This is proven by an analogous calculation. The expression for $\Pr_{\mathbf{y}\sim\mathcal{U}_\tau(u)}[\mathbf{y} = v]$ is given by dropping the $\frac{1}{n^d}$ factor from (10) and then plugging in the expressions obtained in (11) and (12). The quantity $\Pr_{\boldsymbol{H}\sim\mathbb{H}(u),\, \mathbf{y}\sim\mathcal{U}_{\boldsymbol{H},\tau}(u)}[\mathbf{y} = v]$ is precisely $\Pr[\mathcal{E}_v \mid \mathcal{E}_u]$, and an expression for this is obtained by (14) and plugging in the expressions obtained in (15) and (16). Thus, (1) and (3) are equivalent and this completes the proof. □

### A.2 Influence and Persistence Proofs

**Claim A.1.** *If $\widetilde{I}_f > 9\sqrt{d}$, then $\widetilde{I}_f^- > \sqrt{d}$.*

*Proof.* Theorem 9.1 of [KMS18] asserts that for any $\boldsymbol{H}$, if $I_{\boldsymbol{H}} > 6\sqrt{d}$, then $I_{\boldsymbol{H}}^- > I_{\boldsymbol{H}}/3$. (This holds for any Boolean hypercube function.) If $\widetilde{I}_f > 9\sqrt{d}$, then by Claim 3.14, $\mathbb{E}_{\boldsymbol{H}}[I_{\boldsymbol{H}}] > 9\sqrt{d}$. Hence,

$$\begin{aligned} 9\sqrt{d} < \mathbb{E}_{\boldsymbol{H}}[I_{\boldsymbol{H}}] &= \Pr[I_{\boldsymbol{H}} \le 6\sqrt{d}]\ \mathbb{E}_{\boldsymbol{H}}[I_{\boldsymbol{H}} | I_{\boldsymbol{H}} \le 6\sqrt{d}] \ +\ \Pr[I_{\boldsymbol{H}} > 6\sqrt{d}]\ \mathbb{E}_{\boldsymbol{H}}[I_{\boldsymbol{H}} | I_{\boldsymbol{H}} > 6\sqrt{d}] \\ &< 6\sqrt{d} + \Pr[I_{\boldsymbol{H}} > 6\sqrt{d}]\mathbb{E}_H[3I_{\boldsymbol{H}}^- | I_{\boldsymbol{H}} > 6\sqrt{d}] \le 6\sqrt{d} + 3\mathbb{E}_{\boldsymbol{H}}[I_{\boldsymbol{H}}^-] \end{aligned}$$

Hence, $\mathbb{E}_{\boldsymbol{H}}[I_{\boldsymbol{H}}^-] > \sqrt{d}$. By Claim 3.14, $\widetilde{I}_f^- > \sqrt{d}$. □

**Claim A.2.** *If $\widetilde{I}_f \le 9\sqrt{d}$, then the fraction of vertices that are not $(\tau,\beta)$-persistent is at most $C_{per}\frac{\tau}{\beta\sqrt{d}}$ where $C_{per}$ is a universal constant.*

*Proof.* We will analyze the random walk using the distributions described in the first and second bullet point of Fact 3.11 and leverage the analysis that [KMS18] use to prove their Lemma 9.3. Let $\alpha_{up}$ denote the fraction of vertices in the fully augmented hypergrid that are not $(\tau, \beta)$-up-persistent. Using the definition of persistence and Fact 3.11, we have

$$\alpha_{up} \cdot \beta < \Pr_{\mathbf{x} \in_R [n]^d,\, \mathbf{y} \sim \mathcal{U}_\tau(\mathbf{x})} [f(\mathbf{x}) \neq f(\mathbf{z})] = \mathbb{E}_{\boldsymbol{H} \sim \mathbb{H}} \left[ \Pr_{\mathbf{x} \in_R \boldsymbol{H},\, \mathbf{y} \sim \mathcal{U}_{\boldsymbol{H},\tau}(\mathbf{x})} [f(\mathbf{x}) \neq f(\mathbf{z})] \right]. \tag{18}$$

Let $\widehat{\mathcal{U}}_{\boldsymbol{H},\tau}(\mathbf{x})$ denote the same distribution as $\mathcal{U}_{\boldsymbol{H},\tau}(\mathbf{x})$ except with the set $R$ being a uar subset of the 0-coordinates of $\mathbf{x}$. I.e. $\widehat{\mathcal{U}}_{\boldsymbol{H},\tau}(\mathbf{x})$ is the *non-lazy* walk distribution on $\boldsymbol{H}$. Let $\mathbf{x} = \mathbf{x}^0, \mathbf{x}^1, \ldots, \mathbf{x}^\tau = \mathbf{z}$ be the $\tau$ steps taken on the walk sampled by $\mathcal{U}_{\boldsymbol{H},\tau}(\mathbf{x})$ and let $\mathbf{x} = \widehat{\mathbf{x}}^0, \widehat{\mathbf{x}}^1, \ldots, \widehat{\mathbf{x}}^\tau = \mathbf{z}$ be the $\tau$ steps taken on the walk sampled by $\widehat{\mathcal{U}}_{\boldsymbol{H},\tau}(\mathbf{x})$. For a fixed $\boldsymbol{H}$ we have

$$\Pr_{\mathbf{x} \in_R \boldsymbol{H},\, \mathbf{y} \sim \mathcal{U}_{\boldsymbol{H},\tau}(\mathbf{x})} [f(\mathbf{x}) \neq f(\mathbf{z})] \leq \sum_{\ell=0}^{\tau-1} \Pr\left[f(\mathbf{x}^\ell) \neq f(\mathbf{x}^{\ell+1})\right] \leq \sum_{\ell=0}^{\tau-1} \Pr\left[f(\widehat{\mathbf{x}}^\ell) \neq f(\widehat{\mathbf{x}}^{\ell+1})\right]. \tag{19}$$

The first inequality is by a union bound and the second inequality holds because the first walk is lazy and the second is not. More precisely, we can couple the $\tau' \leq \tau$ steps of the lazy-random walk where the point actually moves to the first $\tau'$ steps of the second non-lazy walk, and the remaining $\tau - \tau'$ terms of the non-lazy walk can only increase the RHS.

By Lemma 9.4 of [KMS18], the edge $(\widehat{\mathbf{x}}^\ell, \widehat{\mathbf{x}}^{\ell+1})$ is distributed approximately as a uniform random edge in $\boldsymbol{H}$. In particular, this implies $\Pr\left[f(\widehat{\mathbf{x}}^\ell) \neq f(\widehat{\mathbf{x}}^{\ell+1})\right] \leq C \cdot 2I_{\boldsymbol{H}}/d$ for an absolute constant $C$. (Note $2I_{\boldsymbol{H}}/d$ is the probability of a uniform random edge in $\boldsymbol{H}$ being influential.) Putting (18) and (19) together yields $\alpha_{up} \leq \frac{4C\tau}{d}\mathbb{E}_{\boldsymbol{H}}[I_{\boldsymbol{H}}]$ and an analogous argument gives the same bound for $\alpha_{down}$. Thus, by Claim 3.14 we have $\mathbb{E}_{\boldsymbol{H}}[I_{\boldsymbol{H}}] \leq 9\sqrt{d}$ and the fraction of $(\tau, \beta)$-non-persistent vertices is at most $\frac{72C\tau}{\beta\sqrt{d}}$. Therefore, setting $C_{per} := 72C$ completes the proof. □

## A.3 Typical Points and Reversibility Proofs

**Claim A.3.** *For a $d$-dimensional hypercube $\boldsymbol{H}$ and $c \geq 1$, we have $|\boldsymbol{H}_c| \geq (1 - (\varepsilon/d)^c) \cdot 2^d$.*

*Proof.* Consider a uniform random point $\mathbf{x}$ in the hypercube. The Hamming weight $\|\mathbf{x}\|_1$ is $\sum_{i=1}^d \mathbf{x}_i$, where each $\mathbf{x}_i$ is an iid unbiased Bernoulli. By Hoeffding's theorem, $\Pr[\left|\|\mathbf{x}\|_1 - d/2\right| \geq t] \leq 2\exp(-2t^2/d)$. We set $t = \sqrt{4cd\log(d/\varepsilon)}$. The probability of not being in the $c$-middle layers is at most

$$2\exp(-2t^2/d) = 2\exp(-8c\log(d/\varepsilon)) = 2(\varepsilon/d)^{8c} \leq (\varepsilon/d)^c.$$

Hence, the probability of being in the $c$-middle layers is at least $(1 - (\varepsilon/d)^c)$. □

**Lemma A.4** (Most Points are Typical)**.** *For any $\varepsilon \in (0,1)$ and $c \geq 6$,*

$$\Pr_{\mathbf{x} \in_R [n]^d} [\mathbf{x} \textit{ is } c\textit{-typical}] \geq 1 - (\varepsilon/d)^{c-5}$$

*Proof.* Given $\mathbf{x} \in [n]^d$ and a hypercube $\boldsymbol{H} \ni \mathbf{x}$, let $\chi(\mathbf{x}, \boldsymbol{H}) = \mathbf{1}(\mathbf{x} \in \boldsymbol{H} \setminus \boldsymbol{H}_c)$. By Fact 3.11 and Claim 3.19, we have

$$\mathbb{E}_{\mathbf{x} \in_R [n]^d} \mathbb{E}_{\boldsymbol{H} \sim \mathbb{H}(\mathbf{x})} [\chi(\mathbf{x}, \boldsymbol{H})] = \mathbb{E}_{\boldsymbol{H} \sim \mathbb{H}} \mathbb{E}_{\mathbf{x} \in_R \boldsymbol{H}} [\chi(\mathbf{x}, \boldsymbol{H})] \leq (\varepsilon/d)^c$$

Let us set $q_{\mathbf{x}} := \mathbb{E}_{\boldsymbol{H}\sim\mathbb{H}(\mathbf{x})}[\chi(\mathbf{x},\boldsymbol{H})]$, so $\mathbb{E}_{\mathbf{x}}[q_{\mathbf{x}}] \le (\varepsilon/d)^c$. By Markov's inequality, $\Pr_{\mathbf{x}}[q_{\mathbf{x}} \ge (\varepsilon/d)^5] \le (\varepsilon/d)^{c-5}$. Note that when $q_{\mathbf{x}} < (\varepsilon/d)^5$, $\mathbf{x}$ is $c$-typical. Hence, at least a $(1-(\varepsilon/d)^{c-5})$-fraction of points are $c$-typical. □

**Claim A.5** (Translations of Typical Points). *Suppose $\mathbf{x} \in [n]^d$ is c-typical. Then for a walk length $\tau \le \sqrt{d}$, every point $\mathbf{x}' \in supp(\mathcal{U}_\tau(\mathbf{x})) \cup supp(\mathcal{D}_\tau(\mathbf{x}))$ is $(c+\frac{\tau}{\sqrt{d}})$-typical.*

*Proof.* We prove the claim for $\mathbf{x}' \in \text{supp}(\mathcal{U}_\tau(\mathbf{x}))$. The argument for points in $\text{supp}(\mathcal{D}_\tau(\mathbf{x}))$ is analogous. Let $\boldsymbol{H}$ be any hypercube containing $\mathbf{x}$ and $\mathbf{x}'$ and let $\|\mathbf{x}\|_{\boldsymbol{H}}$, $\|\mathbf{x}'\|_{\boldsymbol{H}}$ denote the Hamming weight of these points in $\boldsymbol{H}$. Observe that $\|\mathbf{x}'\|_{\boldsymbol{H}} \le \|\mathbf{x}\|_{\boldsymbol{H}} + \tau$ and so if $\mathbf{x} \in \boldsymbol{H}_c$, then $\|\mathbf{x}'\|_{\boldsymbol{H}} \le d/2 + \sqrt{cd\log d} + \tau$ and since $\tau \le \sqrt{d}$, we have

$$\sqrt{cd\log d} + \tau \le \sqrt{cd\log d + \tau\sqrt{d}\log d} = \sqrt{\left(c+\frac{\tau}{\sqrt{d}}\right)d\log d}.$$

To see that the first inequality holds, observe that by squaring both sides and rearranging terms, it is equivalent to the inequality

$$\tau^2 + 2\tau\sqrt{cd\log d} \le \tau\sqrt{d}\log d \iff \tau \le \sqrt{d}(\log d - 2\sqrt{c\log d})$$

which clearly holds by our upper bound on $\tau$. Thus, if $\mathbf{x} \in \boldsymbol{H}_c$, then $\mathbf{x}' \in \boldsymbol{H}_{c+\frac{\tau}{\sqrt{d}}}$. Therefore, the number of hypercubes $\boldsymbol{H}$ for which $\mathbf{x}' \in \boldsymbol{H}_{c+\frac{\tau}{\sqrt{d}}}$ is at least the number of hypercubes $\boldsymbol{H}$ for which $\mathbf{x} \in \boldsymbol{H}_c$. Therefore $\mathbf{x}'$ is $(c+\frac{\tau}{\sqrt{d}})$-typical. □

**Lemma A.6** (Reversibility Lemma). *For any points $\mathbf{x} \prec \mathbf{x}' \in [n]^d$ and walk length $\ell \le \sqrt{d}/\log^5(d/\varepsilon)$, we have*

$$p_{\mathbf{x},\ell}(\mathbf{x}') = (1 \pm \log^{-3} d)p_{\mathbf{x}',\ell}(\mathbf{x}).$$

*Proof.* If $t := \|\mathbf{x}-\mathbf{x}'\|_0 > \ell$, then $p_{\mathbf{x},\ell}(\mathbf{x}') = p_{\mathbf{x}',\ell}(\mathbf{x}) = 0$. So assume $t \le \ell$. Fix any $\boldsymbol{H}$ containing $\mathbf{x}$ and $\mathbf{x}'$ such that $\mathbf{x},\mathbf{x}' \in \boldsymbol{H}_{100}$ and let $x$ and $x'$ denote the corresponding hypercube (bit) representations of $\mathbf{x},\mathbf{x}'$ in $\boldsymbol{H}$. Let $p_{x,\ell}(x') = \Pr_{z\sim\mathcal{U}_{\boldsymbol{H},\ell}(x)}[z=x']$ and $p_{x',\ell}(x) = \Pr_{z\sim\mathcal{D}_{\boldsymbol{H},\ell}(x')}[z=x]$. It suffices to show that $p_{x,\ell}(x') = (1\pm\log^{-3} d)p_{x',\ell}(x)$.

Let $S$ be the set of $t$ coordinates where $x$ and $x'$ differ. Let $Z(x)$ be the set of zero coordinates of the point $x$; analogously, define $Z(x')$. Recall that the directed upward walk making $\ell$ steps might not flip $\ell$ coordinates. The process (recall Definition 3.10) picks a uar set $R$ of $\ell$ coordinates, and only flips the zero bits in $x$ among $R$. Hence, an $\ell$-length walk leads from $x$ to $x'$ iff $R \cap Z(x) = S$.

Let the Hamming weight of $x$ be represented as $d/2 + e_x$, where $e_x$ denotes the "excess". Since $x$ is in the 100-middle layers, $|e_x| \le \sqrt{400d\log(d/\varepsilon)}$.

The sets $R$ that lead from $x$ to $y$ can be constructed by picking any $\ell - t$ coordinates in $\overline{Z(x)}$ and choosing all remaining coordinates to be $S$. Hence,

$$p_{x,\ell}(x') = \frac{\binom{d/2+e_x}{\ell-t}}{\binom{d}{\ell}}$$

Analogously, consider the downward $\ell$ step walks from $x'$. This walk leads to $x$ iff $R \cap \overline{Z(x')} = S$. The sets $R$ that lead from $y$ to $x$ can be constructed by picking any $\ell - t$ coordinates in $Z(x')$ and choosing all remaining coordinates to be $S$. The size of $Z(x')$ is precisely $|Z(x)| - t = d/2 - e_x - t$. Hence,

$$p_{x',\ell}(x) = \frac{\binom{d/2-e_x-t}{\ell-t}}{\binom{d}{\ell}}$$

Taking the ratio,

$$\begin{aligned}\frac{p_{x,\ell}(x')}{p_{x',\ell}(x)} &= \frac{\binom{d/2+e_x}{\ell-t}}{\binom{d/2-e_x-t}{\ell-t}} = \frac{\prod_{i=0}^{\ell-t-1}(d/2+e_x-i)}{\prod_{i=0}^{\ell-t-1}(d/2-e_x-t-i)} = \prod_{i=0}^{\ell-t-1}\frac{d/2+e_x-i}{d/2-e_x-t-i} \\ &= \prod_{i=0}^{\ell-t-1}\left(1+\frac{2e_x+t}{d/2-e_x-t-i}\right)\end{aligned}$$

Recall that $|e_x| \leq \sqrt{400d\log(d/\varepsilon)}$, $t \leq \ell < \sqrt{d}/\log^5(d/\varepsilon)$. For convenience, let $b := \sqrt{400d\log(d/\varepsilon)}$. So $2e_x + t \leq 3b$. Also, $d/2 - e_x - t - i \geq d/3$ for all $i < \ell$. Applying these bounds,

$$\frac{p_{x,\ell}(x')}{p_{x',\ell}(x)} \leq \prod_{i=0}^{\ell-1}\left(1+\frac{3b}{d/3}\right) \leq \exp\left(\frac{9\ell b}{d}\right) = \exp\left(\frac{\sqrt{d}\cdot\sqrt{400d\log(d/\varepsilon)}}{d\log^5(d/\varepsilon)}\right) \leq 1+\log^{-3} d$$

An analogous calculation proves that $\frac{p_{x,\ell}(x')}{p_{x',\ell}(x)} \geq 1 - \log^{-3} d$. $\square$



ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il