On Mappings on the Hypercube with Small Average Stretch

Lucas Boczkowski  
CNRs, IRIF Université Paris 7
lucasboczko@gmail.com

Igor Shinkar  
Simon Fraser University
ishinkar@sfu.ca

Abstract

Let \( A \subseteq \{0, 1\}^n \) be a set of size \( 2^n - 1 \), and let \( \phi: \{0, 1\}^{n-1} \to A \) be a bijection. We define the average stretch of \( \phi \) as
\[
\text{avgStretch}(\phi) = \mathbb{E}[\text{dist}(\phi(x), \phi(x'))],
\]
where the expectation is taken over uniformly random \( x, x' \in \{0, 1\}^{n-1} \) that differ in exactly one coordinate.

In this paper we continue the line of research studying mappings on the discrete hypercube with small average stretch. We prove the following results.

- For any set \( A \subseteq \{0, 1\}^n \) of density \( 1/2 \) there exists a bijection \( \phi_A: \{0, 1\}^{n-1} \to A \) such that \( \text{avgStretch}(\phi_A) = O(\sqrt{n}) \).
- For \( n = 3^k \) let \( A_{\text{rec-maj}} = \{ x \in \{0, 1\}^n : \text{rec-maj}(x) = 1 \} \), where \( \text{rec-maj}: \{0, 1\}^n \to \{0, 1\} \) is the function recursive majority of 3’s. There exists a bijection \( \phi_{\text{rec-maj}}: \{0, 1\}^{n-1} \to A_{\text{rec-maj}} \) such that \( \text{avgStretch}(\phi_{\text{rec-maj}}) = O(1) \).
- Let \( A_{\text{tribes}} = \{ x \in \{0, 1\}^n : \text{tribes}(x) = 1 \} \). There exists a bijection \( \phi_{\text{tribes}}: \{0, 1\}^{n-1} \to A_{\text{tribes}} \) such that \( \text{avgStretch}(\phi_{\text{tribes}}) = O(\log(n)) \).

These results answer the questions raised by Benjamini, Cohen, and Shinkar (Isr. J. Math 2016).
Contents

1 Introduction ............................................. 3
   1.1 A uniform upper bound on the average stretch .................................... 4
   1.2 Bounds on the average stretch for specific sets .................................... 4
      1.2.1 Recursive majority of 3’s ...................................................... 5
      1.2.2 The tribes function .......................................................... 5
   1.3 Roadmap ................................................. 6

2 Proof of Theorem 1.3 ...................................... 6
   2.1 Upper bound on the average transportation distance using stable marriage .... 6
   2.2 Proof of Theorem 1.3 using transportation theory .................................... 7

3 Average stretch for recursive majority of 3’s ........................................... 8

4 Average stretch for tribes ...................................... 13

5 Concluding remarks and open problems .................................................. 16
1 Introduction

In this paper we continue the line of research from [BCS14, RS18, JS21] studying geometric similarities between different subsets of the hypercube $\mathcal{H}_n = \{0,1\}^n$. Given a set $A \subseteq \mathcal{H}_n$ of size $|A| = 2^n - 1$ and a bijection $\phi: \mathcal{H}_{n-1} \rightarrow A$, we define the average stretch of $\phi$ as

$$\text{avgStretch}(\phi) = \mathbb{E}_{x \sim x' \in \mathcal{H}_{n-1}}[\text{dist}(\phi(x), \phi(x'))],$$

where $\text{dist}(x, y)$ is defined as the number of coordinates $i \in [n]$ such that $x_i \neq y_i$, and the expectation is taken over a uniformly random $x, x' \in \mathcal{H}_{n-1}$ that differ in exactly one coordinate.\footnote{Note that any $C$-Lipschitz function $\phi: \mathcal{H}_{n-1} \rightarrow A$ satisfies $\text{avgStretch}(\phi) \leq C$. That is, the notion of average stretch is a relaxation of the Lipschitz property.}

The origin of this notion is motivated by the study of the complexity of distributions [GGN10, Vio12, LV12]. In this line of research, given a distribution $\mathcal{D}$ on $\mathcal{H}_n$, the goal is to find a mapping $h: \mathcal{H}_m \rightarrow \mathcal{H}_n$ such that if $U_m$ is the uniform distribution over $\mathcal{H}_m$, then $h(U_m)$ is (close to) the distribution $\mathcal{D}$, and each output bit $h_i$ of the function $h$ is computable efficiently (e.g., computable in $AC_0$, i.e., by polynomial size circuits of constant depth).

Motivated by the goal of proving lower bounds for sampling from the uniform distribution on some set $A \subseteq \mathcal{H}_n$, Lovett and Viola [LV12] suggested the restricted problem of proving that no bijection from $\mathcal{H}_{n-1}$ to $A$ can be computed in $AC_0$. Toward this goal they noted that it suffices to prove that any such bijection requires large average stretch. Indeed, by the structural results of [Has86, Bop97, LMN93] it is known that any such mapping $\phi$ that is computable by a polynomial size circuit of depth $d$ has $\text{avgStretch}(\phi) < \log(n)^{O(d)}$, and hence proving that any bijection requires super-polylogarithmic average stretch implies that it cannot be computed in $AC_0$. Proving a lower bound for sampling using this approach remains an open problem.

Studying this problem, [BCS14] have shown that if $n$ is odd, and $A_{maj} \subseteq \mathcal{H}_n$ is the hamming ball of density $1/2$, i.e. $A_{maj} = \{x \in \mathcal{H}_n : \sum_i x_i > n/2\}$, then there is an $O(1)$-bi-Lipschitz mapping from $\mathcal{H}_{n-1}$ to $A_{maj}$, thus suggesting that proving a lower bound for a bijection from $\mathcal{H}_{n-1}$ to $A_{maj}$ requires new ideas beyond the sensitivity-based structural results of [Has86, Bop97, LMN93] mentioned above. In [RS18] it has been shown that if a subset $A_{\text{rand}}$ of density $1/2$ is chosen uniformly at random, then with high probability there is a bijection $\phi: \mathcal{H}_{n-1} \rightarrow A_{\text{rand}}$ with $\text{avgStretch}(\phi) = O(1)$. This result has recently been improved by Johnston and Scott [JS21], who showed that for a random set $A_{\text{rand}} \subseteq \mathcal{H}_n$ of density $1/2$ there exists an $O(1)$-Lipschitz bijection from $\mathcal{H}_{n-1}$ to $A_{\text{rand}}$ with high probability.

The following problem was posed in [BCS14], and repeated in [RS18, JS21].

**Problem 1.1.** Exhibit a subset $A \subseteq \mathcal{H}_n$ of density $1/2$ such that any bijection $\phi: \mathcal{H}_{n-1} \rightarrow A$ has $\text{avgStretch}(\phi) = \omega(1)$, or prove that no such subset exists.\footnote{Throughout the paper, the density of a set $A \subseteq \mathcal{H}_n$ is defined as $\mu_n(A) = |A|/2^n$.}

**Remark** Note that it is easy to construct a set of density $1/2$ such that any bijection $\phi: \mathcal{H}_{n-1} \rightarrow A$ must have a worst case stretch at least $n/2$. For example, for odd $n$ consider the set $A = \{y \in \mathcal{H}_n : n/2 < \sum_i y_i < n\} \cup \{0^n\}$. Then any bijection $\phi: \mathcal{H}_{n-1} \rightarrow A$ must map some point $x \in \mathcal{H}_{n-1}$ to $0^n$, while all neighbours $x'$ of $x$ are mapped to some $\phi(x')$ with weight at least $n/2$. Hence, the worst case stretch of $\phi$ is at least $n/2$. In contrast, Problem 1.1 does not seem to have a non-trivial solution.
To rephrase Problem 1.1, we are interested in determining a tight upper bound on the \( \text{avgStretch} \) that holds uniformly for all sets \( A \subseteq \mathcal{H}_n \) of density 1/2. Note that since the diameter of \( \mathcal{H}_n \) is \( n \), for any set \( A \subseteq \mathcal{H}_n \) of density 1/2 and any bijection \( \phi: \mathcal{H}_{n-1} \to A \) it holds that \( \text{avgStretch}(\phi) \leq n \). It is natural to ask how tight this bound is, i.e., whether there exists \( A \subseteq \mathcal{H}_n \) of density 1/2 such that any bijection \( \phi: \mathcal{H}_{n-1} \to A \) requires linear average stretch. It is consistent with our current knowledge (though hard to believe) that for any set \( A \setminus \mathcal{H}_n \) such that \( \text{avgStretch}(A) \) is for the set \( A \) that holds uniformly for all sets \( \mathcal{H}_n \).

Most of the research on metric embedding, we are aware of, focuses on worst case stretch. For a survey on metric embeddings of finite spaces see [Lin02]. There has been a lot of research on the question of embedding into the Boolean cube. For example, see [AB07, HLN87] for work on embeddings between random subsets of the Boolean cube, and [Gra88] for isometric embeddings of arbitrary graphs into the Boolean cube.

### 1.1 A uniform upper bound on the average stretch

We prove a non-trivial uniform upper bound on the average stretch of a mapping \( \phi: \mathcal{H}_{n-1} \to A \) that applies to all sets \( A \subseteq \mathcal{H}_n \) of density 1/2.

**Theorem 1.2.** For any set \( A \subseteq \mathcal{H}_n \) of density \( \mu_n(A) = 1/2 \), there exists a bijection \( \phi: \mathcal{H}_{n-1} \to A \) such that \( \text{avgStretch}(\phi) = O(\sqrt{n}) \).

Toward this goal we prove a stronger result bounding the average transportation distance between two arbitrary sets of density 1/2. Specifically, we prove the following theorem.

**Theorem 1.3.** For any two sets \( A, B \subseteq \mathcal{H}_n \) of density \( \mu_n(A) = \mu_n(B) = 1/2 \), there exists a bijection \( \phi: A \to B \) such that \( \mathbb{E}[\text{dist}(x, \phi(x))] \leq \sqrt{2n} \).

Note that Theorem 1.2 follows immediately from Theorem 1.3 by the following simple argument.

**Proposition 1.4.** Fix a bijection \( \phi: \mathcal{H}_{n-1} \to A \). Then \( \text{avgStretch}(\phi) \leq 2\mathbb{E}_{x \in \mathcal{H}_{n-1}}[\text{dist}(x, \phi(x))] + 1 \).

**Proof.** Using the triangle inequality we have

\[
\text{avgStretch}(\phi) = \mathbb{E}_{x \in \mathcal{H}_{n-1}}[\text{dist}(\phi(x), \phi(x + e_i))] \\
\leq \mathbb{E}[\text{dist}(x, \phi(x))] + \text{dist}(x, x + e_i) + \mathbb{E}[\text{dist}(x + e_i, \phi(x + e_i))] \\
= \mathbb{E}[\text{dist}(x, \phi(x))] + 1 + \mathbb{E}[\text{dist}(x + e_i, \phi(x + e_i))] \\
= 2\mathbb{E}[\text{dist}(x, \phi(x))] + 1,
\]

as required. \( \square \)

### 1.2 Bounds on the average stretch for specific sets

Next, we study two specific subsets of \( \mathcal{H}_n \) defined by Boolean functions commonly studied in the field “Analysis of Boolean functions” [O’D14]. Specifically, we study two monotone noise-sensitive functions: the recursive majority of 3’s, and the tribes function.
It was suggested in [BCS14] that the set of ones of these functions $A_f = f^{-1}(1)$ may be such that any mapping $\phi : \mathcal{H}_{n-1} \rightarrow A_f$ requires large $\text{avgStretch}$. We show that for the recursive majority function there is a mapping $\phi_{\text{rec-maj}} : \mathcal{H}_{n-1} \rightarrow \text{rec-maj}^{-1}(1)$ with $\text{avgStretch}(\phi_{\text{rec-maj}}) = O(1)$. For the tribes function we show a mapping $\phi_{\text{tribes}} : \mathcal{H}_{n-1} \rightarrow \text{tribes}^{-1}(1)$ with $\text{avgStretch}(\phi_{\text{tribes}}) = O(\log(n))$. Below we formally define the functions, and discuss our results.

### 1.2.1 Recursive majority of 3’s

The recursive majority of 3’s function is defined as follows.

**Definition 1.5.** Let $k \in \mathbb{N}$ be a positive integer. Define the function recursive majority of 3’s $\text{rec-maj}_k : \mathcal{H}_{3^k} \rightarrow \{0,1\}$ as follows.

- For $k = 1$ the function $\text{rec-maj}_1$ is the majority function on the 3 input bits.
- For $k > 1$ the function $\text{rec-maj}_k : \mathcal{H}_{3^k} \rightarrow \{0,1\}$ is defined recursively as follows. For each $x \in \mathcal{H}_{3^k}$ write $x = x^{(1)} \circ x^{(2)} \circ x^{(3)}$, where each $x^{(r)} \in \mathcal{H}_{3^{k-1}}$ for each $r \in [3]$. Then, $\text{rec-maj}_k(x) = \text{maj}((\text{rec-maj}_{k-1}(x^{(1)}), \text{rec-maj}_{k-1}(x^{(2)}), \text{rec-maj}_{k-1}(x^{(3)})))$.

Note that $\text{rec-maj}_k(x) = 1 - \text{rec-maj}_k(1 - x)$ for all $x \in \mathcal{H}_n$, and hence the density of the set $A_{\text{rec-maj}_k} = \{x \in \mathcal{H}_n : \text{rec-maj}_k(x) = 1\}$ is $\mu_n(A_{\text{rec-maj}_k}) = 1/2$. We prove the following result regarding the set $A_{\text{rec-maj}_k}$.

**Theorem 1.6.** Let $k$ be a positive integer, let $n = 3^k$, and let $A_{\text{rec-maj}_k} = \{x \in \mathcal{H}_n : \text{rec-maj}_k(x) = 1\}$. There exists a mapping $\phi_{\text{rec-maj}_k} : \mathcal{H}_{n-1} \rightarrow A_{\text{rec-maj}_k}$ such that $\text{avgStretch}(\phi_{\text{rec-maj}_k}) \leq 20$.

### 1.2.2 The tribes function

The tribes function is defined as follows.

**Definition 1.7.** Let $s, w \in \mathbb{N}$ be two positive integers, and let $n = sw$. The function tribes : $\mathcal{H}_n \rightarrow \{0,1\}$ is defined as a DNF consisting of $s$ disjoint clauses of width $w$.

$$\text{tribes}(x_1, x_2, \ldots, x_w; \ldots; x_{(s-1)w+1} \ldots x_{sw}) = \bigvee_{i=1}^{s} (x_{(i-1)w+1} \land x_{(i-1)w+2} \land \cdots \land x_{iw}).$$

That is, the function tribes partitions $n = sw$ inputs into $s$ disjoint “tribes” each of size $w$, and returns 1 if and only if at least one of the tribes “votes” 1 unanimously.

It is clear that $\Pr_{x \in \mathcal{H}_n}[\text{tribes}(x) = 1] = 1 - (1 - 2^{-w})^s$. The interesting settings of parameters $w$ and $s$ are such that the function is close to balanced, i.e., this probability is close to 1/2. Given $w \in \mathbb{N}$, let $s = s_w = \ln(2)2^w + \Theta(1)$ be the largest integer such that $1 - (1 - 2^{-w})^s \leq 1/2$. For such choice of the parameters we have $\Pr_{x \in \mathcal{H}_n}[\text{tribes}(x) = 1] = \frac{1}{2} - O\left(\frac{\log(n)}{n}\right)$ (see, e.g., [O’D14, Section 4.2]).

Consider the set $A_{\text{tribes}} = \{x \in \mathcal{H}_n : \text{tribes}(x) = 1\}$. Since the density of $A_{\text{tribes}}$ is not necessarily equal to 1/2, we cannot talk about a bijection from $\mathcal{H}_{n-1}$ to $A_{\text{tribes}}$. In order to overcome this technical issue, let $A^*_{\text{tribes}}$ be an arbitrary superset of $A_{\text{tribes}}$ of density 1/2. We prove that there is a mapping $\phi_{\text{tribes}}$ from $\mathcal{H}_{n-1}$ to $A^*_{\text{tribes}}$ with average stretch $\text{avgStretch}(\phi_{\text{tribes}}) = O(\log(n))$. In fact, we prove a stronger result, namely that the average transportation distance of $\phi_{\text{tribes}}$ is $O(\log(n))$. 


Theorem 1.8. Let $w$ be a positive integer, and let $s$ be the largest integer such that $1 - (1 - 2^{-w})^s \leq 1/2$. For $n = s \cdot w$ let $\mathcal{H}_n \to \{0, 1\}$ be defined as a DNF consisting of $s$ disjoint clauses of width $w$. Let $A_{\text{tribes}} = \{x \in \mathcal{H}_n : \text{tribes}(x) = 1\}$, and let $A_{\text{tribes}}^* \subseteq \mathcal{H}_n$ be an arbitrary superset of $A_{\text{tribes}}$ of density $\mu_n(A_{\text{tribes}}) = 1/2$. Then, there exists a bijection $\phi_{\text{tribes}} : \mathcal{H}_{n-1} \to A_{\text{tribes}}^*$ such that $\mathbb{E}[\text{dist}(x, \phi_{\text{tribes}}(x))] = O(\log(n))$. In particular, $\text{avgStretch}(\phi_{\text{tribes}}) = O(\log(n))$.

1.3 Roadmap

The rest of the paper is organized as follows. We prove Theorem 1.3 in Section 2. In Section 3 we prove Theorem 1.6, and in Section 4 we prove Theorem 1.8.

2 Proof of Theorem 1.3

We provide two different proofs of Theorem 1.3. The first proof, in Section 2.1 shows a slightly weaker bound of $O\left(\sqrt{n \ln(n)}\right)$ on the average stretch using the Gale-Shapley result on the stable marriage problem. The idea of using the stable marriage problem was suggested in [BCS14], and we implement this approach. Then, in Section 2.2, we show the bound of $O(\sqrt{n})$ by relating the average stretch of a mapping between two sets to known estimates on the Wasserstein distance on the hypercube.

2.1 Upper bound on the average transportation distance using stable marriage

Recall the Gale-Shapley theorem on the stable marriage problem. In the stable marriage problem we are given two sets of elements $A$ and $B$ each of size $N$. For each element $a \in A$ (resp. $b \in B$) we have a ranking of the elements of $B$ (resp. $A$) given as an bijection $rk_a : A \to [N]$ ($rk_b : B \to [N]$) representing the preferences of each $a$ (resp. $b$). A matching (or a bijection) $\phi : A \to B$ is said to be unstable if there are $a, a' \in A$, and $b, b' \in B$ such that $\phi(a) = b'$, $\phi(a') = b$, but $rk_a(b) < rk_a(b')$, and $rk_b(a) < rk_b(a')$; that is, both $a$ and $b$ would prefer to be mapped to each other over their mappings given by $\phi$. We say that a matching $\phi : A \to B$ is stable otherwise.

Theorem 2.1 (Gale-Shapley theorem). For any two sets $A$ and $B$ of equal size and any choice of rankings for each $a \in A$ and $b \in B$ there exists a stable matching $m : A \to B$.

For the proof of Theorem 1.3 the sets $A$ and $B$ are subsets of $\mathcal{H}_n$ of density $1/2$. We define the preferences between points based on the distances between them in $\mathcal{H}_n$. That is, for each $a \in A$ we have $rk_a(b) < rk_a(b')$ if and only if $\text{dist}(a, b) < \text{dist}(a, b')$ with ties broken arbitrarily. Similarly, for each $b \in B$ we have $rk_b(a) < rk_b(a')$ if and only if $\text{dist}(a, b) < \text{dist}(a', b)$ with ties broken arbitrarily.

Let $\phi : A \to B$ be a bijection. We show that if $\mathbb{E}_{x \in A}[\text{dist}(x, \phi(x))] > 3k$ for $k = \left\lceil \sqrt{n \ln(n)} \right\rceil$, then $\phi$ is not a stable matching. Consider the set

$$F := \{x \in A \mid \text{dist}(x, \phi(x)) \geq k\}.$$

Note that since the diameter of $\mathcal{H}_n$ is $n$, and $\mathbb{E}_{x \in A}[\text{dist}(x, \phi(x))] > 3k$, it follows that $\mu_n(F) > \frac{k}{n}$. Indeed, we have $3k < \mathbb{E}_{x \in A}[\text{dist}(x, \phi(x))] \leq n \cdot \frac{\mu_n(F)}{\mu_n(A)} + k \cdot (1 - \frac{\mu_n(F)}{\mu_n(A)}) \leq n \cdot \frac{\mu_n(F)}{\mu_n(A)} + k$, and thus $\mu_n(F) > \frac{2k}{n} \cdot \mu_n(A)$. Next, we use Talagrand’s concentration inequality.
**Theorem 2.2** ([Tal95, Proposition 2.1.1]). Let $k \leq n$ be two positive integers, and let $F \subseteq \mathcal{H}_n$. Let $F_{\geq k} = \{ x \in \mathcal{H}_n : \text{dist}(x,y) \geq k \; \forall y \in F \}$ be the set of all $x \in \mathcal{H}_n$ whose distance from $F$ is at least $k$. Then $\mu_n(F_{\geq k}) \leq e^{-k^2/n}/\mu_n(F)$.

By Theorem 2.2 we have $\mu_n(F_{\geq k}) \leq e^{-k^2/n}/\mu_n(F)$, and hence, for $k = \left\lceil \sqrt{n \ln(n)} \right\rceil$ it holds that

$$\mu_n(F_{\geq k}) \leq e^{-\ln(n)}/\mu_n(F) \leq (1/n)/(k/n) = 1/k.$$ 

In particular, since $\mu_n(\phi(F)) = \mu_n(F) > k/n > 1/k \geq \mu_n(F_{\geq k})$, there is some $b \in \phi(F)$ that does not belong to $F_{\geq k}$. That is, there is some $a \in F$ and $b \in \phi(F)$ such that $\text{dist}(a,b) < k$. On the other hand, for $a' = \phi^{-1}(b)$, by the definition of $F$ we have $\text{dist}(a,\phi(a)) \geq k$ and $\text{dist}(a',b = \phi(a')) \geq k$, and hence $\phi$ is not stable, as $a$ and $b$ prefer to be mapped to each other over their current matching. Therefore, in a stable matching $E_{x \in \mathcal{A}}[\text{dist}(x,\phi(x))] \leq 3\left\lceil \sqrt{n \ln(n)} \right\rceil$, and by the Gale-Shapley theorem such a matching does, indeed, exist.

### 2.2 Proof of Theorem 1.3 using transportation theory

Next we prove Theorem 1.3, by relating our problem to a known estimate on the Wasserstein distance between two measures on the hypercube. Recall that the $\ell_1$-Wasserstein distance between two measures $\mu$ and $\nu$ on $\mathcal{H}_n$ is defined as

$$W_1(\mu, \nu) = \inf_{q} \sum_{x,y} \text{dist}(x,y)q(x,y),$$

where the infimum is taken over all couplings $q$ of $\mu$ and $\nu$, i.e., $\sum_y q(x,y) = \mu(x)$ and $\sum_x q(x',y) = \nu(y)$ for all $x, y \in \mathcal{H}_n$. That is, we consider an optimal coupling $q$ of $\mu$ and $\nu$ minimizing $E_{(x,y) \sim q}[\text{dist}(x,y)]$, the expected distance between $x$ and $y$, where $x$ is distributed according to $\mu$ and $y$ is distributed according to $\nu$.

We prove the theorem using the following two claims.

**Claim 2.3.** Let $\mu_A$ and $\mu_B$ be uniform measures over the sets $A$ and $B$ respectively. Then, there exists a bijection $\phi$ from $A$ to $B$ such that $E[\text{dist}(x,\phi(x))] = W_1(\mu_A, \mu_B)$.

**Claim 2.4.** Let $\mu_A$ and $\mu_B$ be uniform measures over the sets $A$ and $B$ respectively. Then $W_1(\mu_A, \mu_B) \leq \sqrt{2n}$.

**Proof of Claim 2.3.** Observe that any bijection $\phi$ from $A$ to $B$ naturally defines a coupling $q$ of $\mu_A$ and $\mu_B$, defined as

$$q(x,y) = \begin{cases} \frac{1}{|A|} & \text{if } x \in A \text{ and } y = \phi(x), \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, $W_1(\mu_A, \mu_B) \leq E_{x \in A}[\text{dist}(x,\phi(x))]$.

For the other direction note that in the definition of $W_1$ we are looking for the infimum of the linear function $L(q) = \sum_{(x,y) \in A \times B} \text{dist}(x,y)q(x,y)$, where the infimum is taken over the **Birkhoff polytope** of all $n \times n$ doubly stochastic matrices. By the Birkhoff-von Neumann theorem [Bir46, vN53, Kôn36] this polytope is the convex hull whose extremal points are precisely the permutation matrices. The optimum is obtained on such an extremal point, and hence there exists a bijection $\phi$ from $A$ to $B$ such that $W_1(\mu_A, \mu_B) = E[\text{dist}(x,\phi(x))]$. 

\[ \square \]
Proof of Claim 2.4. The proof of the claim follows rather directly from the techniques in transportation theory (see [RS13, Section 3.4]). Specifically, using Definition 3.4.2 and combining Proposition 3.4.1, Equation 3.4.42, and Proposition 3.4.3, where $\mathcal{X} = \{0, 1\}$, and $\mu$ is the uniform distribution on $\mathcal{X}$ we have the following theorem.

**Theorem 2.5.** Let $\nu$ be an arbitrary distribution on the discrete hypercube $\mathcal{H}_n$, and let $\mu_n$ be the uniform distribution on $\mathcal{H}_n$. Then

$$W_1(\nu, \mu_n) \leq \sqrt{\frac{1}{2} n \cdot D(\nu \parallel \mu_n)},$$

where $D(\nu \parallel \mu)$ is the Kullback-Leibler divergence defined as $D(\nu \parallel \mu) = \sum x \nu(x) \log(\frac{\nu(x)}{\mu(x)})$.

In particular, by letting $\nu = \mu_A$ be the uniform distribution over the set $A$ of cardinality $2^{n-1}$, we have $D(\mu_A \parallel \mu_n) = \sum_{x \in A} \mu_A(x) \log(\frac{\mu_A(x)}{\mu_n(x)}) = \sum_{x \in A} \frac{1}{|A|} \log(2) = 1$, and hence $W_1(\mu_n, \mu_A) \leq \sqrt{\frac{1}{2} n \cdot D(\mu_n \parallel \nu)} = \sqrt{n/2}$. Analogously, we have $W_1(\mu_n, \mu_B) \leq \sqrt{n/2}$. Therefore, by the triangle inequality, we conclude that $W_1(\mu_A, \mu_B) \leq W_1(\mu_A, \mu_n) + W_1(\mu_n, \mu_B) \leq \sqrt{2n}$, as required. \qed

This completes the proof of Theorem 1.3.

## 3 Average stretch for recursive majority of 3’s

In this section we prove Theorem 1.6, showing a mapping from $\mathcal{H}_n$ to $A_{\text{rec-maj}_k}$ with constant average stretch. The key step in the proof is the following lemma.

**Lemma 3.1.** Let $k$ be a positive integer, and let $n = 3^k$. There exists $f_k : \mathcal{H}_n \to A_{\text{rec-maj}_k}$ satisfying the following properties.

1. $f_k(x) = x$ for all $x \in A_{\text{rec-maj}_k}$.
2. For each $x \in A_{\text{rec-maj}_k}$ there is a unique $z \in Z_{\text{rec-maj}_k} := \mathcal{H}_n \setminus A_{\text{rec-maj}_k}$ such that $f_k(z) = x$.
3. For every $i \in [n]$ we have $E_{x \in \mathcal{H}_n}[\text{dist}(f_k(x), f_k(x + e_i))] \leq 10$.

We postpone the proof of Lemma 3.1 for now, and show how it implies Theorem 1.6.

**Proof of Theorem 1.6.** Let $f_k$ be the mapping from Lemma 3.1. Define $\psi_0, \psi_1 : \mathcal{H}_{n-1} \to A_{\text{rec-maj}_k}$ as $\psi_0(x) = f_k(x \circ 0)$, where $x \circ b \in \mathcal{H}_n$ is the string obtained from $x$ by appending to it $b$ as the $n$’th coordinate.

The mappings $\psi_0, \psi_1$ naturally induce a bipartite graph $G = (V, E)$, where $V = \mathcal{H}_{n-1} \cup A_{\text{rec-maj}_k}$ and $E = \{(x, \psi_0(x)) : x \in \mathcal{H}_{n-1}, b \in \{0, 1\}\}$, possibly containing parallel edges. Note that by the first two items of Lemma 3.1 the graph $G$ is 2-regular. Indeed, for each $x \in \mathcal{H}_n$ the neighbours of $x$ are $N(x) = \{\psi_0(x) = f_k(x \circ 0), \psi_1(x) = f_k(x \circ 1)\}$, and for each $y \in A_{\text{rec-maj}_k}$ there is a unique $x \in A_{\text{rec-maj}_k}$ and a unique $z \in Z_{\text{rec-maj}_k}$ such that $f_k(x) = f_k(z) = 1$, and hence $N(y) = \{x[^{1,...,n-1}], z[^{1,...,n-1}]\}$.

Since the bipartite graph $G$ is 2-regular, it has a perfect matching. Let $\phi$ be the bijection from $\mathcal{H}_{n-1}$ to $A_{\text{rec-maj}_k}$ induced by a perfect matching in $G$, and for each $x \in \mathcal{H}_n$ let $b_x \in \mathcal{H}_n$ be such...
that \( \phi(x) = \psi_{b_x}(x) \). We claim that \( \text{avgStretch}(\phi) = O(1) \). Let \( x \sim x' \) be uniformly random vertices in \( \mathcal{H}_{n-1} \) that differ in exactly one coordinate, and let \( r \in \{0, 1\} \) be uniformly random. Then

\[
\mathbb{E}[\text{dist}(\phi(x), \phi(x'))] = \mathbb{E}[\text{dist}(f_k(x \circ b_x), f_k(x' \circ b_{x'}))]
\leq \mathbb{E}[\text{dist}(f_k(x \circ b_x), f_k(x \circ r)) + \text{dist}(f_k(x \circ r), f_k(x' \circ r))]
\leq \mathbb{E}[\text{dist}(f_k(x' \circ r), f_k(x' \circ b_{x'}))].
\]

For the first term, since \( r \) is equal to \( b_x \) with probability 1/2 by Lemma 3.1 Item 3 we get that \( \mathbb{E}[\text{dist}(f_k(x \circ b_x), f_k(x \circ r))] \leq 5 \). Analogously the third term is bounded by 5. In the second term we consider the expected distance between \( f(\cdot) \) applied on inputs that differ in a random coordinate \( i \in [n-1] \), which is at most 10, again, by Lemma 3.1 Item 3. Therefore \( \mathbb{E}[\text{dist}(\phi(x), \phi(x'))] \leq 20 \). \( \square \)

We return to the proof of Lemma 3.1.

**Proof of Lemma 3.1.** Define \( f_k: \mathcal{H}_n \to A_{\text{rec-maj}}_k \) by induction on \( k \). For \( k = 1 \) define \( f_1 \) as

\[
\begin{align*}
000 &\mapsto 110 \\
100 &\mapsto 101 \\
010 &\mapsto 011 \\
001 &\mapsto 111 \\
x &\mapsto x \quad \text{for all } x \in \{110, 101, 011, 111\}.
\end{align*}
\]

That is, \( f_1 \) acts as the identity map for all \( x \in A_{\text{rec-maj}}_1 \), and maps all inputs in \( Z_{\text{rec-maj}}_1 \) to \( A_{\text{rec-maj}}_1 \) in a one-to-one way. Note that \( f_1 \) is a non-decreasing mapping, i.e., \( (f_1(x))_i \geq x_i \) for all \( x \in \mathcal{H}_3 \) and \( i \in [3] \).

For \( k > 1 \) define \( f_k \) recursively using \( f_{k-1} \) as follows. For each \( r \in [3] \), let \( T_r = [(r-1) \cdot 3^{k-1} + 1, \ldots, r \cdot 3^{k-1}] \) be the \( r \)’th third of the interval \( [3^k] \). For \( x \in \mathcal{H}_{3^k} \), write \( x = x^{(1)} \circ x^{(2)} \circ x^{(3)} \), where \( x^{(r)} = x^{T_r} \in \mathcal{H}_{3^{k-1}} \) is the \( r \)’th third of \( x \). Let \( y = (y_1, y_2, y_3) \in \{0,1\}^3 \) be defined as \( y_r = \text{rec-maj}_{k-1}(x^{(r)}) \), and let \( w = (w_1, w_2, w_3) = f_1(y) \in \{0,1\}^3 \). Define

\[
(f_k(x^{(r)}))^{(r)} = \begin{cases} f_{k-1}(x^{(r)}) & \text{if } w_r \neq y_r, \\ x^{(r)} & \text{otherwise.} \end{cases}
\]

Finally, the mapping \( f_k \) is defined as

\[
f_k(x) = f_{k-1}^{(1)}(x^{(1)}) \circ f_{k-1}^{(2)}(x^{(2)}) \circ f_{k-1}^{(3)}(x^{(3)}).
\]

That is, if \( \text{rec-maj}_k(x) = 1 \) then \( w = y \), and hence \( f_k(x) = x \), and otherwise, \( f_{k-1}^{(r)}(x^{(r)}) \neq x^{(r)} \) for all \( r \in [3] \) where \( y_r = 0 \) and \( w_r = 1 \).

Next we prove that \( f_k \) satisfies the properties stated in Lemma 3.1.

1. It is clear from the definition that if \( \text{rec-maj}_k(x) = 1 \), then \( w = y \), and hence \( f_k(x) = x \).

2. Next, we prove by induction on \( k \) that the restriction of \( f_k \) to \( Z_{\text{rec-maj}}_k \) induces a bijection. For \( k = 1 \) the statement clearly holds. For \( k > 2 \) suppose that the restriction of \( f_{k-1} \) to \( Z_{\text{rec-maj}}_{k-1} \) induces a bijection. We show that for every \( x \in A_{\text{rec-maj}}_k \) the mapping \( f_k \) has a preimage of \( x \) in \( Z_{\text{rec-maj}}_k \). Write \( x = x^{(1)} \circ x^{(2)} \circ x^{(3)} \), where \( x^{(r)} = x^{T_r} \in \mathcal{H}_{3^{k-1}} \) is the \( r \)’th
third of $x$. Let $w = (w_1, w_2, w_3)$ be defined as $w_r = \text{rec-maj}_{k-1}(x^{(r)})$. Since $x \in A_{\text{rec-maj}_k}$ it follows that $w \in \{110, 101, 011, 111\}$. Let $y = (y_1, y_2, y_3) \in Z_{\text{rec-maj}_1}$ such that $f_1(y) = w$.

For each $r \in [3]$ such that $w_r = 1$ and $y_r = 0$ it must be the case that $x^{(r)} \in A_{\text{rec-maj}_{k-1}}$, and hence, by the induction hypothesis, there is some $z^{(r)} \in Z_{\text{rec-maj}_{k-1}}$ such that $f_{k-1}(z^{(r)}) = x^{(r)}$.

For each $r \in [3]$ such that $y_r = w_r$ define $z^{(r)} = x^{(r)}$. Since $y = (y_1, y_2, y_3) \in Z_{\text{rec-maj}_1}$, it follows that $z = z^{(1)} \circ z^{(2)} \circ z^{(3)} \in Z_{\text{rec-maj}_k}$. It is immediate by the construction that, indeed, $f_k(z) = x$.

3. Fix $i \in [3^k]$. In order to prove $E[\text{dist}(f_k(x), f_k(x + e_i))] = O(1)$ consider the following events.

$$
E_1 = \{\text{rec-maj}_k(x) = 1 \Leftrightarrow \text{rec-maj}_k(x + e_i)\},
\quad E_2 = \{\text{rec-maj}_k(x) = 0, \text{rec-maj}_k(x + e_i) = 1\},
\quad E_3 = \{\text{rec-maj}_k(x) = 1, \text{rec-maj}_k(x + e_i) = 0\},
\quad E_4 = \{\text{rec-maj}_k(x) = 0 \Leftrightarrow \text{rec-maj}_k(x + e_i)\}.
$$

Then $E[\text{dist}(f_k(x), f_k(x + e_i))] = \sum_{j=1,2,3,4} E[\text{dist}(f_k(x), f_k(x + e_i))|E_j] \cdot P[E_j]$. The following three claims prove an upper bound on $E[\text{dist}(f_k(x), f_k(x + e_i))]$.

**Claim 3.2.** $E[\text{dist}(f_k(x), f_k(x + e_i))|E_1] = 1$.

**Claim 3.3.** $E[\text{dist}(f_k(x), f_k(x + e_i))|E_2] \leq 2 \cdot 1.5^k$.

**Claim 3.4.** $E[\text{dist}(f_k(x), f_k(x + e_i))|E_4] \cdot P[E_4] \leq 8$.

By symmetry, it is clear that $E[\text{dist}(f_k(x), f_k(x + e_i))|E_2] = E[\text{dist}(f_k(x), f_k(x + e_i))|E_3]$. Note also that $P[E_1] < 0.5$ and $P[E_2 \cup E_3] = 2^{-k}$. Therefore, the claims above imply that

$$
E[\text{dist}(f_k(x), f_k(x + e_i))] = \sum_{j=1,2,3,4} E[\text{dist}(f_k(x), f_k(x + e_i))|E_j] \cdot P[E_i] \leq 1.05 + 2 \cdot 1.5^k \cdot 2^{-k} + 8 \leq 10,
$$

which completes the proof of Lemma 3.1.

Next we prove the above claims.

**Proof of Claim 3.2.** If $E_1$ holds then $\text{dist}(f_k(x), f_k(x + e_i)) = \text{dist}(x, x + e_i) = 1$.

**Proof of Claim 3.3.** We prove first that

$$
E[\text{dist}(x, f_k(x))|\text{rec-maj}_k(x) = 0] = 1.5^k.
$$

The proof is by induction on $k$. For $k = 1$ we have $E[\text{dist}(x, f_1(x))|\text{rec-maj}_k(x) = 0] = 1.5$ as there are two inputs $x \in Z_{\text{rec-maj}_k}$ with $\text{dist}(x, f_1(x)) = 1$ and two $x$‘s in $Z_{\text{rec-maj}_k}$ with $\text{dist}(x, f_1(x)) = 2$. For $k > 1$ suppose that $E[\text{dist}(x, f_{k-1}(x))|\text{rec-maj}_{k-1}(x)] = 1.5^{k-1}$. Write each $x \in H_{3^k}$ as $x = x^{(1)} \circ x^{(2)} \circ x^{(3)}$, where $x^{(r)} = x_{r(i)} \in H_{3^{k-1}}$ is the $r$‘th third of $x$, and let $y = (y_1, y_2, y_3)$ be

4Indeed, note that $P[E_2 \cup E_3] = P[\text{rec-maj}_k(x) \neq \text{rec-maj}_k(x + e_i)]$, and suppose for concreteness that $i = 1$. We claim that $P[\text{rec-maj}_k(x) \neq \text{rec-maj}_k(x + e_i)] = 2^{-k}$, which can be seen by induction on $k$ using the recurrence $P[\text{rec-maj}_k(x) \neq \text{rec-maj}_k(x + e_i)] = P[\text{rec-maj}_{k-1}(x^{(2)}) \neq \text{rec-maj}_{k-1}(x^{(2)})] \cdot P[\text{rec-maj}_{k-1}(x^{(1)}) \neq \text{rec-maj}_{k-1}(x^{(1)}) + e_i)] = (1/2) \cdot P[\text{rec-maj}_{k-1}(x^{(1)}) \neq \text{rec-maj}_{k-1}(x^{(1)} + e_i)] = (1/2) \cdot 2^{-k} = 2^{-k}$. 


defined as \( y_r = \text{rec-maj}_{k-1}(x^{(r)}) \). Since \( E_{x \in H_{k-1}}[\text{rec-maj}_{k-1}(x)] = 0.5 \), it follows that for a random \( z \in Z_{\text{rec-maj}_k} \) each \( y \in \{001, 100, 010, 000\} \) happens with the same probability 1/4, and hence, using the induction hypothesis we get

\[
E[\text{dist}(x, f_k(x)) | \text{rec-maj}_k(x) = 0] = \frac{\mathbb{P}[y \in \{100, 010\} | \text{rec-maj}_k(x) = 0]}{4} \times 1.5^{k-1} + \frac{\mathbb{P}[y \in \{000, 001\} | \text{rec-maj}_k(x) = 0]}{4} \times 2 \times 1.5^{k-1} = 1.5^k,
\]

which proves Eq. (1).

Next we prove that

\[
E[\text{dist}(x, f_k(x)) | E_2] \leq \sum_{j=0}^{k-1} 1.5^j = 2 \cdot (1.5^k - 1).
\]

Note that Eq. (2) proves Claim 3.3. Indeed, if \( E_2 \) holds then using the triangle inequality we have

\[
\text{dist}(f_k(x), f_k(x+e_i)) \leq \text{dist}(f_k(x), x) + \text{dist}(x, x+e_i) + \text{dist}(x+e_i, f_k(x+e_i)) = \text{dist}(f_k(x), x) + 1,
\]

and hence

\[
E[\text{dist}(f_k(x), x) | E_2] + 1 \leq 2 \cdot (1.5^k - 1) + 1 < 2 \cdot 1.5^k,
\]
as required.

We prove Eq. (2) by induction on \( k \). For \( k = 1 \) Eq. (2) clearly holds. For the induction step let \( k > 1 \). As in the definition of \( f_k \) write each \( x \in H_{3k} \) as \( x = x^{(1)} \circ x^{(2)} \circ x^{(3)} \), where \( x^{(r)} = x_{T_r} \) is the \( r \)th third of \( x \), and let \( y = (y_1, y_2, y_3) \) be defined as \( y_r = \text{rec-maj}_{k-1}(x^{(r)}) \).

Let us suppose for concreteness that \( i \in T_1 \). (The cases of \( i \in T_2 \) and \( i \in T_3 \) are handled similarly.) Note that if \( \text{rec-maj}_k(x) = 0 \), \( \text{rec-maj}_k(x+e_i) = 1 \), and \( i \in T_1 \), then \( y \in \{010, 001\} \). We consider each case separately.

1. Suppose that \( y = 010 \). Then \( w = f(y) = 011 \), and hence \( f(x) \) differs from \( x \) only in \( T_3 \).

Taking the expectation over \( x \) such that \( \text{rec-maj}_k(x) = 0 \) and \( \text{rec-maj}_k(x+e_i) = 1 \) by Eq. (1) we get \( E[\text{dist}(x, f(x)) | E_2, y = 010] = E[\text{dist}(f_{k-1}(x^{(3)}), x^{(3)}) | \text{rec-maj}_{k-1}(x^{(3)}) = 0] = 1.5^{k-1} \).

2. If \( y = 001 \), then \( w = f_1(y) = 111 \), and \( f(x) \) differs from \( x \) only in \( T_1 \cup T_2 \). Then

\[
E[\text{dist}(x, f(x)) | E_2, y = 001] = E[\text{dist}(f_{k-1}(x^{(1)}), x^{(1)}) | E_2, y = 001]
\]

\[
+ E[\text{dist}(f_{k-1}(x^{(2)}), x^{(2)}) | E_2, y = 001].
\]

Denoting by \( E'_2 \) the event that \( \text{rec-maj}_{k-1}(x^{(1)}) = 0, \text{rec-maj}_{k-1}(x^{(1)}+e_i) = 1 \) (i.e., the analogue of the event \( E_2 \) applied on \( \text{rec-maj}_{k-1} \)), we note that

\[
E[\text{dist}(f_{k-1}(x^{(1)}), x^{(1)}) | E_2, y = 001] = E[\text{dist}(f_{k-1}(x^{(1)}), x^{(1)}) | E'_2],
\]

which is upper bounded by \( \sum_{j=0}^{k-2} 1.5^j \) using the induction hypothesis. For the second term we have

\[
E[\text{dist}(f_{k-1}(x^{(2)}), x^{(2)}) | E_2, y = 001] = E[\text{dist}(f_{k-1}(x^{(2)}), x^{(2)}) | \text{rec-maj}_{k-1}(x^{(2)}) = 0],
\]

\footnote{Note that Eq. (2) can be thought of Eq. (1) conditioned on the event \( \text{rec-maj}_k(x+e_i) = 1 \), which happens with probability only \( 2^{-k} \). A naive application of Markov's inequality would only say that \( E[\text{dist}(x, f_k(x)) | E_2] \leq 1.5^k \cdot 2^k \), which would not suffice for us. Eq. (2) says that the expected distance is comparable to \( 1.5^k \) even when conditioning on this small event.}
which is at most $1.5^{k-1}$ using Eq. (1). Therefore, for $y = 001$ we have

$$
\mathbb{E}[\text{dist}(x, f(x)) \mid E_2, y = 001] \leq \sum_{j=0}^{k-2} 1.5^j + 1.5^{k-1}.
$$

Using the two cases for $y$ we get

$$
\mathbb{E}[\text{dist}(x, f_k(x)) \mid E_2] = \mathbb{E}[\text{dist}(x, f_k(x)) \mid E_2, y = 010] \cdot \mathbb{P}[y = 010 \mid E_2] + \mathbb{E}[\text{dist}(x, f_k(x)) \mid E_2, y = 001] \cdot \mathbb{P}[y = 001 \mid E_2]
$$

$$
\leq \sum_{j=0}^{k-1} 1.5^j.
$$

This proves Eq. (2) for the case where $i \in T_1$. The other two cases are handled similarly. This completes the proof of Claim 3.3.

\[\square\]

**Proof of Claim 3.4.** For a coordinate $i \in [n]$ and for $0 \leq j \leq k$ let $r = r_i(j) \in \mathbb{N}$ be such that $i \in [(r-1) \cdot 3^j + 1, \ldots, r \cdot 3^j]$, and denote the corresponding interval by $T_i(j) = [(r-1) \cdot 3^j + 1, \ldots, r \cdot 3^j].$ These are the coordinates used in the recursive definition of $\text{rec-maj}_k$ by the instance of $\text{rec-maj}_j$ that depends on the $i$’th coordinate.

For $x \in H_n$ and $x' = x + e_i$, define $\nu(x)$ as

$$
\nu(x) = \begin{cases} 
\min \{ j \in [k] : \text{rec-maj}_j(x_{T_i(j)}) = \text{rec-maj}_j(x'_{T_i(j)}) \} & \text{if } \text{rec-maj}_k(x) = \text{rec-maj}_k(x'), \\
 k + 1 & \text{if } \text{rec-maj}_k(x) \neq \text{rec-maj}_k(x').
\end{cases}
$$

That is, in the ternary tree defined by the computation of $\text{rec-maj}_k$, $\nu(x)$ is the lowest $j$ on the path from the $i$’th coordinate to the root where the computation of $x$ is equal to the computation of $x + e_i$. Note that if $x$ is chosen uniformly from $H_n$, then

$$
\mathbb{P}[\nu = j] = \begin{cases} 
2^{-j} & \text{if } j \in [k], \\
2^{-k} & \text{if } j = k + 1.
\end{cases}
$$

Below we show that by conditioning on $E_4$ and on the value of $\nu$ we get

$$
\mathbb{E}[\text{dist}(f_k(x), f_k(x + e_i)) \mid E_4, \nu = j] \leq 4 \cdot 1.5^j.
$$

Indeed, suppose that $E_4$ holds. Assume without loss of generality that $x_i = 0$, and let $x' = x + e_i$. Note that $f_k(x)$ and $f_k(x')$ differ only on the coordinates in the interval $T_i(\nu)$. Let $w = x_{T_i(\nu)}$, and define $y = (y_1, y_2, y_3) \in \{0, 1\}^3$ as $y_r = \text{rec-maj}_{\nu-1}(w^{(r)})$ for each $r \in [3]$, where $w^{(r)}$ is the $r$’th third of $w$. Similarly, let $w' = x'_{T_i(\nu)}$, and let $y' = (y'_1, y'_2, y'_3) \in \{0, 1\}^3$ be defined as $y'_r = \text{rec-maj}_{\nu-1}(w'^{(r)})$ for each $r \in [3]$. This implies that

$$
\mathbb{E}[\text{dist}(f_k(x), f_k(x + e_i)) \mid E_4] = \mathbb{E}[\text{dist}(f_\nu(w), f_\nu(w')) \mid E_4].
$$

\[\text{Footnote 5}\text{For example, for } j = 0 \text{ we have } T_i(0) = \{i\} \text{ For } j = 1 \text{ if } i \equiv 1 \pmod{3} \text{ then } T_i(1) = [i, i + 1, i + 2]. \text{ For } j = k - 1 \text{ the interval } T_i(k - 1) \text{ is one of the intervals } T_1, T_2, T_3. \text{ For } j = k \text{ we have } T_i(k) = [1, \ldots, 3^k].\]
Furthermore, if \( \text{rec-maj}_\nu(x_{T_i(\nu)}) = 1 \) (and \( \text{rec-maj}_\nu(x'_{T_i(\nu)}) = 1 \)), then \( f_k(x)_{T_i(\nu)} = x_{T_i(\nu)} \), and thus \( \text{dist}(f_k(x), f_k(x')) = 1 \).

Next we consider the case of \( \text{rec-maj}_k(x_{T_i(\nu)}) = 0 \) (and \( \text{rec-maj}_k(x'_{T_i(\nu)}) = 0 \)). Since \( x_i = 0 \) and \( x' = x + e_i \), it must be that \( y = 000 \) and \( y' \) is a unit vector. Suppose first that \( y' = 100 \), i.e., the coordinate \( i \) belongs to the first third of \( T_i(\nu) \). Write \( w = w^{(1)} \circ w^{(2)} \circ w^{(3)} \), where each \( w^{(r)} \) is one third of \( w \). Analogously, write \( w' = w'^{(1)} \circ w'^{(2)} \circ w'^{(3)} \), where each \( w'^{(r)} \) one third of \( w' \). Then, since \( w' = w + e_i \) we have

\[
\mathbb{E}[\text{dist}(f_j(w), w)|E_4, \nu = j] = \mathbb{E}[\text{dist}(f_j(w^{(1)}), w^{(1)}]|\text{rec-maj}_{j-1}(w^{(1)}) = 0, \text{rec-maj}_{j-1}(w^{(1)} + e_i) = 1] \\
+ \mathbb{E}[\text{dist}(f_j(w^{(2)}), w^{(2)})|\text{rec-maj}_{j-1}(w^{(2)}) = 0] \\
\leq 2 \cdot (1.5^j - 1) + 1.5^j = 3 \cdot 1.5^j - 2,
\]

where the last inequality is by Eq. (1) and Eq. (2). Similarly,

\[
\mathbb{E}[\text{dist}(f_j(w'), w'|E_4, \nu = j] = \mathbb{E}[\text{dist}(f_j(w'^{(3)}), w'^{(3)})|\text{rec-maj}_{j-1}(w'^{(3)}) = 0] \leq 1.5^j - 1,
\]

where the last inequality is by Eq. (1). Therefore,

\[
\mathbb{E}[\text{dist}(f_j(w), f_j(w'))|E_4, \nu = j] < 4 \cdot 1.5^j - 1.
\]

The cases of \( y = 010 \) and \( 001 \) are handled similarly, and it is straightforward to verify that in these cases we also get the bound of \( 4 \cdot 1.5^j - 1 \).

By combining Eq. (3) with Eq. (4) it follows that

\[
\mathbb{E}[\text{dist}(f_k(x), f_k(x + e_i))|E_4] \cdot \mathbb{P}[E_4] = \sum_{j=1}^{k} \mathbb{E}[\text{dist}(f_k(x), f_k(x + e_i))|E_4, \nu = j] \cdot \mathbb{P}[\nu = j|E_4] \cdot \mathbb{P}[E_4] \\
\leq \sum_{j=1}^{k} 4 \cdot 1.5^j - 1 \cdot \mathbb{P}[\nu = j] \\
\leq \sum_{j=1}^{k} 4 \cdot 1.5^j - 1 \cdot 2^{-j} \leq 8.
\]

This completes the proof of Claim 3.4. \( \square \)

## 4 Average stretch for tribes

In this section we prove Theorem 1.8, showing a mapping from \( \mathcal{H}_n \) to \( A^*_\text{tribes} \) with \( O(\log(n)) \) average stretch. Let \( \mu^1\text{tribes} \) be the uniform distribution on \( A^*_\text{tribes} \), and let \( \mu^0\text{tribes} \) be the uniform distribution on \( Z^0\text{tribes} = \mathcal{H}_n \setminus A^*_\text{tribes} \). The proof consists of the following two claims.

**Claim 4.1.** For \( \mu^1\text{tribes} \) and \( \mu^0\text{tribes} \) as above it holds that

\[
W_1(\mu^0\text{tribes}, \mu^1\text{tribes}) = O(\log(n)).
\]

Next, let \( A^*\text{tribes} \subseteq \mathcal{H}_n \) be an arbitrary superset of \( A^*_\text{tribes} \) of density 1/2, and let \( \mu^*\text{tribes} \) be the uniform distribution on \( A^*\text{tribes} \).
Claim 4.2. Consider $\mathcal{H}_{n-1}$ as $\{x \in \mathcal{H}_n : x_n = 0\}$, and let $\mu_{n-1}$ be the uniform measure on $\mathcal{H}_{n-1}$. Then,

$$W_1(\mu_{n-1}, \mu^*_\text{tribes}) \leq W_1(\mu^0_\text{tribes}, \mu^1_\text{tribes}) + O(\log(n)).$$

By combining Claim 4.1 and Claim 4.2 we get that the average transportation distance between $\mathcal{H}_{n-1}$ and $A^*_\text{tribes}$ is $W_1(\mu_{n-1}, \mu^*_\text{tribes}) = O(\log(n))$. By Claim 2.3 it follows that there exists $\phi^*_\text{tribes} : \mathcal{H}_{n-1} \rightarrow A^*_\text{tribes}$ such that $\mathbb{E}[\text{dist}(x, \phi^*_\text{tribes}(x))] = O(\log(n))$, and using Proposition 1.4 we conclude that $\text{avgStretch}(\phi^*_\text{tribes}) = O(\log(n))$. This completes the proof of Theorem 1.8.

Below we prove Claim 4.1 and Claim 4.2.

Proof of Claim 4.1. Let $\mathcal{D} = \mathcal{D}_w$ the uniform distribution over $\{0,1\}^w \setminus \{1^w\}$, let $p = 2^{-w}$, and denote by $\mathcal{L} = \mathcal{L}_{w,s}$ the binomial distribution $\text{Bin}(p, s)$ conditioned on the outcome being positive. That is,

$$\mathbb{P}[\mathcal{L} = \ell] = \frac{\binom{s}{\ell}p^\ell(1-p)^{s-\ell}}{\sum_{j=1}^{s} \binom{s}{j}p^j(1-p)^{s-j}} \quad \forall \ell \in \{1, \ldots, s\}.$$ 

Note that $\mu^0_\text{tribes}$ is equal to the product distribution $\mathcal{D}^s$. Note also that in order to sample from the distribution $\mu^1_\text{tribes}$, we can first sample $\mathcal{L} \in \{1, \ldots, s\}$, then choose $\mathcal{L}$ random tribes that vote unanimously 1, and for the remaining $s - \mathcal{L}$ tribes sample their values in this tribe according to $\mathcal{D}$. We define a coupling $q_\text{tribes}$ between $\mu^0_\text{tribes}$ and $\mu^1_\text{tribes}$ as follows. First sample $x$ according to $\mu^0_\text{tribes}$. Then, sample $\mathcal{L} \in \{1, \ldots, s\}$, choose $\mathcal{L}$ random tribes $T \subseteq [s]$ uniformly, and let $S = \{(t - 1)w + j : t \in T, j \in [w]\}$ be all the coordinates participating in all tribes in $T$. Define $y \in \mathcal{H}_n$ as $y_i = 1$ for all $i \in S$, and $y_i = x_i$ for all $i \in [n] \setminus S$. It is clear that $y$ is distributed according to $\mu^1_\text{tribes}$, and hence $q^*_\text{tribes}$ is indeed a coupling between $\mu^0_\text{tribes}$ and $\mu^1_\text{tribes}$.

We next show that $\mathbb{E}_{(x,y) \sim q_\text{tribes}}[\text{dist}(x,y)] = O(\log(n))$. We have $\mathbb{E}_{(x,y) \sim q_\text{tribes}}[\text{dist}(x,y)] \leq \mathbb{E}[\mathcal{L} : w]$, and by the choice of parameters, we have $w \leq \log(n)$ and $\mathbb{E}[\mathcal{L}] = \frac{\mathbb{E}[\text{Bin}(2^{-w}, s)]}{1 - \mathbb{P}[\text{Bin}(2^{-w}, s) = 0]} = \frac{s2^{-w}}{1 - 2^{-w}}$. By the choice of $s \leq \ln(2)2^w + O(1)$ it follows that $\mathbb{E}[\mathcal{L}] = O(1)$, and hence

$$W_1(\mu^0_\text{tribes}, \mu^1_\text{tribes}) \leq \mathbb{E}_{(x,y) \sim q_\text{tribes}}[\text{dist}(x,y)] \leq \mathbb{E}[\mathcal{L} : w] = O(\log(n)).$$

This completes the proof of Claim 4.1.

Proof of Claim 4.2. We start by showing that

$$W_1(\mu_n, \mu^1_\text{tribes}) \leq W_1(\mu^0_\text{tribes}, \mu^1_\text{tribes}),$$

where $\mu_n$ is the uniform measure on $\mathcal{H}_n$. Indeed, let $q_\text{tribes}$ be a coupling between $\mu^0_\text{tribes}$ and $\mu^1_\text{tribes}$. Define a coupling $q_n$ between $\mu_n$ and $\mu^1_\text{tribes}$ as

$$q_n(x,y) = \begin{cases} 
\frac{|Z_{\text{tribes}}|}{2^w} \cdot q_\text{tribes}(x,y) & \text{if } x \in Z_{\text{tribes}} \text{ and } y \in A_{\text{tribes}}, \\
1/2^w & \text{if } x = y \in A_{\text{tribes}}, \\
0 & \text{otherwise}.
\end{cases}$$

It is straightforward to verify that $q_n$ is indeed a coupling between $\mu_n$ and $\mu^1_\text{tribes}$. Letting $q_\text{tribes}$ be a coupling for which $\mathbb{E}_{(x,y) \sim q_\text{tribes}}[\text{dist}(x,y)] = W_1(\mu^0_\text{tribes}, \mu^1_\text{tribes})$ we get

$$W_1(\mu_n, \mu^1_\text{tribes}) \leq \sum_{x \in \mathcal{H}_n} \text{dist}(x,y) q_n(x,y).$$

14
= \sum_{x \in \text{tribes}} \sum_{y \in \text{tribes}} \text{dist}(x, y) q_n(x, y) + \sum_{x \in \text{tribes}} \sum_{y \in \text{tribes}} \text{dist}(x, y) q_n(x, y)

= \frac{|Z_{\text{tribes}}|}{2^n} \mathbb{E}_{(x, y) \sim q_n} [\text{dist}(x, y)] + \sum_{x \in \text{tribes}} \text{dist}(x, x) q_n(x, x)

= \frac{|Z_{\text{tribes}}|}{2^n} \cdot W_1(\mu_{\text{tribes}}^0, \mu_{\text{tribes}}^1) \leq W_1(\mu_{\text{tribes}}^0, \mu_{\text{tribes}}^1),

which proves Eq. (5).

Next, we show that

\[ W_1(\mu_{n-1}, \mu_{\text{tribes}}^1) \leq W_1(\mu_{n-1}, \mu_{\text{tribes}}^0) + 1. \]  \hspace{1cm} (6)

Indeed, let \( q_n \) be a coupling between \( \mu_n \) and \( \mu_{\text{tribes}}^0 \) minimizing \( \sum_{(x, y) \in \mathcal{H}_n \times \text{tribes}} \text{dist}(x, y) q_n(x, y) \).

Define a coupling \( q_{n-1} \) between \( \mu_{n-1} \) and \( \mu_{\text{tribes}}^0 \) as

\[ q_{n-1}(x, y) = q_n(x, y) + q_n(x + e_i, y) \quad \forall x \in \mathcal{H}_{n-1} \text{ and } y \in \text{tribes}. \]

It is clear that \( q_{n-1} \) is a coupling between \( \mu_{n-1} \) and \( \mu_{\text{tribes}}^0 \). Next we prove Eq. (6).

\[
W_1(\mu_{n-1}, \mu_{\text{tribes}}^1) \leq \sum_{x \in \mathcal{H}_{n-1}} \sum_{y \in \text{tribes}} \text{dist}(x, y) q_{n-1}(x, y)
\]

\[
= \sum_{x \in \mathcal{H}_{n-1}} \sum_{y \in \text{tribes}} \text{dist}(x, y) q_n(x, y) + \sum_{x \in \mathcal{H}_{n-1}} \sum_{y \in \text{tribes}} \text{dist}(x, y) q_n(x + e_i, y)
\]

\[
\leq \sum_{x \in \mathcal{H}_{n-1}} \sum_{y \in \text{tribes}} \text{dist}(x, y) q_n(x, y) + \sum_{x \in \mathcal{H}_{n-1}} \sum_{y \in \text{tribes}} (\text{dist}(x + e_i, y) + 1) q_n(x + e_i, y)
\]

\[
= \sum_{x \in \mathcal{H}_n} \sum_{y \in \text{tribes}} \text{dist}(x, y) q_n(x, y) + \sum_{x \in \mathcal{H}_{n-1}} \sum_{y \in \text{tribes}} q_n(x + e_i, y)
\]

\[
\leq W_1(\mu_{n-1}, \mu_{\text{tribes}}^0) + 1,
\]

which proves Eq. (6).

Next, we show that

\[ W_1(\mu_{n-1}, \mu_{\text{tribes}}^0) \leq W_1(\mu_{n-1}, \mu_{\text{tribes}}^0) + O(\log(n)). \]  \hspace{1cm} (7)

In order to prove Eq. (7), let \( \delta = \frac{1}{2} - \frac{|A_{\text{tribes}}^*|}{2^n} \). By the discussion in Section 1.2.2 we have \( \delta = O\left(\frac{\log(n)}{n}\right) \). Then \( |A_{\text{tribes}}^* \setminus \text{tribes}| = \delta \cdot 2^n \). Let \( q_{n-1} \) be a coupling between \( \mu_{n-1} \) and \( \mu_{\text{tribes}}^0 \) such that \( \mathbb{E}_{(x, y) \sim q_{n-1}} [\text{dist}(x, y)] = W_1(\mu_{n-1}, \mu_{\text{tribes}}^0) \). Define a coupling \( q^* \) between \( \mu_{n-1} \) and \( \mu_{\text{tribes}}^0 \) as

\[
q^*(x, y) = \begin{cases} 
(1 - 2\delta) \cdot q_{n-1}(x, y) & \text{if } x \in \mathcal{H}_{n-1} \text{ and } y \in \text{tribes}, \\
4 \cdot 2^{-2n} & \text{if } x \in \mathcal{H}_{n-1} \text{ and } y \in \text{tribes}^* \setminus \text{tribes}.
\end{cases}
\]

It is straightforward to verify that \( q^* \) is a coupling between \( \mu_{n-1} \) and \( \mu_{\text{tribes}}^0 \). Next we prove Eq. (7).

\[
W_1(\mu_{n-1}, \mu_{\text{tribes}}^0) \leq \sum_{x \in \mathcal{H}_{n-1}} \sum_{y \in \text{tribes}^* \setminus \text{tribes}} \text{dist}(x, y) \cdot q^*(x, y)
\]
\begin{align*}
&= (1 - 2\delta) \sum_{x \in \mathcal{H}_{n-1}} \text{dist}(x, y) q_{n-1}(x, y) + \sum_{x \in \mathcal{H}_{n-1}} \text{dist}(x, y) \cdot 4 \cdot 2^{-2n} \\
&\leq (1 - 2\delta) \cdot W_1(\mu_{n-1}, \mu_{\text{tribes}}) + 2\delta \cdot \max_{x \in \mathcal{H}_{n-1}} (\text{dist}(x, y)).
\end{align*}

Eq. (7) follows from the fact that \(\max(\text{dist}(x, y)) \leq n\) and \(\delta = O\left(\frac{\log(n)}{n}\right)\).

By combining Eqs. (5) to (7) we get \(W_1(\mu_{n-1}, \mu_{\text{tribes}}^*) \leq W_1(\mu_{\text{tribes}}^0, \mu_{\text{tribes}}^1) + O(\log(n))\). \qed

5 Concluding remarks and open problems

Uniform upper bound on the average stretch. We’ve shown a uniform upper bound of \(O(\sqrt{n})\) on the average transportation distance \(\mathbb{E}[\text{dist}(x, \phi(x))]\) from \(\mathcal{H}_{n-1}\) to any set \(A \subseteq \mathcal{H}_n\) of density \(1/2\), where \(\mathcal{H}_{n-1}\) is treated as \(\{x \in \mathcal{H}_n : x_n = 0\}\). This bound is tight up to a multiplicative constant. Indeed, it is not difficult to see that for any bijection \(\phi \) from \(\mathcal{H}_{n-1}\) to \(A_{\text{maj}} = \{x \in \mathcal{H}_n : \sum_i x_i > n/2\}\) (for odd \(n\)) the average transportation of \(\phi\) is \(\mathbb{E}[\text{dist}(x, \phi(x))] \geq \Omega(\sqrt{n})\).

In contrast, we believe that the upper bound of \(O(\sqrt{n})\) on the average stretch is not tight, and it should be possible to improve it further.

Problem 5.1. Prove/disprove that for any set \(A \subseteq \mathcal{H}_n\) of density \(1/2\) there exists a mapping \(\phi_A : \mathcal{H}_{n-1} \to A\) with \(\text{avgStretch}(\phi) = o(\sqrt{n})\).

The tribes function. Considering our results about the tribes function, we make the following conjecture.

Conjecture 5.2. Let \(w\) be a positive integer, and let \(s\) be the largest integer such that \(1 - (1 - 2^{-w})^s \leq 1/2\). For \(n = s \cdot w\) let \(\text{tribes} : \mathcal{H}_n \to \{0, 1\}\) be defined as a DNF consisting of \(s\) disjoint clauses of width \(w\), and let \(A_{\text{tribes}} = \{x \in \mathcal{H}_n : \text{tribes}(x) = 1\}\). There exists \(A_{\text{tribes}}^* \subseteq \mathcal{H}_n\) a superset of \(A_{\text{tribes}}\) of density \(\mu_n(A_{\text{tribes}}^*) = 1/2\) such that \(W_1(\mu_{n-1}, \mu_{\text{tribes}}^*) = O(1)\), where \(\mu_{\text{tribes}}^*\) is the uniform distribution on \(A_{\text{tribes}}^*\).

As a first step toward the conjecture we propose the following strengthening of Claim 4.1.

Problem 5.3. Let \(\mu_{\text{tribes}}^1\) be the uniform distribution on \(A_{\text{tribes}}^*\), and let \(\mu_{\text{tribes}}^0\) be the uniform distribution on \(Z_{\text{tribes}} = \mathcal{H}_n \setminus A_{\text{tribes}}\). Is it true that \(W_1(\mu_{\text{tribes}}^0, \mu_{\text{tribes}}^1) = O(1)\)?

A candidate set that requires large average stretch. We propose a candidate set \(A^*\) for which we hope that any mapping from \(\mathcal{H}_{n-1}\) to \(A^*\) requires large average stretch. The set is defined as follows. Let \(k^* \in [n]\) be the maximal \(k\) such that \(\binom{n}{\leq k} = \sum_{j=0}^{k} \binom{n}{j} \leq 2^{n-2}\). Let \(B_{1/4}^0 = \{x \in \mathcal{H}_n : \sum_{i \in [n]} x_i \leq k\}\) and \(B_{1/4}^1 = \{x \in \mathcal{H}_n : \sum_{i \in [n]} x_i \geq n - k\}\) be two (disjoint) antipodal balls of radius \(k^*\), and let \(C \subseteq \mathcal{H}_n \setminus (B_{1/4}^0 \cup B_{1/4}^1)\) be an arbitrary set of size \(|C| = 2^{n-1} - |B_{1/4}^0 \cup B_{1/4}^1|\).

Define \(A^* = B_{1/4}^0 \cup B_{1/4}^1 \cup C\).

Conjecture 5.4. There is no bijection \(\phi^* : \mathcal{H}_{n-1} \to A^*\) with \(\text{avgStretch}(\phi^*) = O(1)\).
Acknowledgements

We are grateful to the anonymous referees for carefully reading the paper, and for their very helpful comments that greatly improved the presentation.

References

[AB07] O. Angel and I. Benjamini. A phase transition for the metric distortion of percolation on the hypercube. *Combinatorica*, 27(6):645–658, 2007.

[BCS14] I. Benjamini, G. Cohen, and I. Shinkar. Bi-Lipschitz bijection between the boolean cube and the Hamming ball. In *FOCS*, pages 81–89, 2014.

[Bir46] G. Birkhoff. Three observations on linear algebra. *Univ. Nac. Tacuman, Rev. Ser. A*, 5:147–151, 1946.

[Bop97] R. Boppana. The average sensitivity of bounded-depth circuits. *Information Processing Letters*, 63(5):257–261, 1997.

[GGN10] O. Goldreich, S. Goldwasser, and A. Nussboim. On the implementation of huge random objects. *SIAM Journal on Computing*, 39(7):2761–2822, 2010.

[Gra88] R. L. Graham. Isometric embeddings of graphs. *Selected Topics in Graph Theory*, 3:133–150, 1988.

[Häs86] J. Håstad. Almost optimal lower bounds for small depth circuits. In *Proceedings of the eighteenth annual ACM Symposium on Theory of Computing*, pages 6–20. ACM, 1986.

[HLN87] J. Håstad, T. Leighton, and M. Newman. Reconfiguring a hypercube in the presence of faults. In *Proceedings of the nineteenth annual ACM Symposium on Theory of Computing*, pages 274–284, 1987.

[JS21] T. Johnston and A. Scott. Lipschitz bijections between boolean functions. *Combinatorics, Probability and Computing*, 30(4):513—525, 2021.

[Kön36] D. König. *Theorie der endlichen und unendlichen Graphen*. Akademische Verlagsgesellschaft, 1936.

[Lin02] N. Linial. Finite metric spaces - combinatorics, geometry and algorithms. In *Proceedings of the International Congress of Mathematicians III*, pages 573–586, 2002.

[LMN93] N. Linial, Y. Mansour, and N. Nisan. Constant depth circuits, fourier transform, and learnability. *Journal of the ACM*, 40(3):607–620, 1993.

[LV12] S. Lovett and E. Viola. Bounded-depth circuits cannot sample good codes. *Computational Complexity*, 21(2):245–266, 2012.

[O’D14] R. O’Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.
[RS13] M. Raginsky and I. Sason. Concentration of measure inequalities in information theory, communications, and coding. *Foundations and Trends in Communications and Information Theory*, 10(1-2):1–246, 2013.

[RS18] S. Rao and I. Shinkar. On Lipschitz bijections between boolean functions. *Combinatorics, Probability and Computing*, 27(3):411–426, 2018.

[Tal95] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathématiques de l’Institut des Hautes Études Scientifiques*, 81(1):73–205, 1995.

[Vio12] E. Viola. The complexity of distributions. *SIAM Journal on Computing*, 41(1):191–218, 2012.

[vN53] J. von Neumann. A certain zero-sum two-person game equivalent to the optimal assignment problem. *Contributions to the Theory of Games II;Annals of Mathematics Studies*, 28:5–12, 1953.