2014

The Information Complexity Of Hamming Distance

E. Blais

Joshua Brody
Swarthmore College, brody@cs.swarthmore.edu

B. Ghazi

Recommended Citation

E. Blais, Joshua Brody, and B. Ghazi. (2014). "The Information Complexity Of Hamming Distance". *Approximation, Randomization, And Combinatorial Optimization: Algorithms And Techniques (APPROX/RANDOM 2014)*. Volume 28, 465-489. DOI: 10.4230/LIPIcs.APPROX-RANDOM.2014.465
https://works.swarthmore.edu/fac-comp-sci/97

The Information Complexity of Hamming Distance

Eric Blais[1], Joshua Brody[2], and Badih Ghazi[1]

1 **MIT, Cambridge, MA, USA**
{eblais|badih}@mit.edu
2 **Swarthmore College, Swarthmore, PA, USA**
joshua.e.brody@gmail.com

Abstract

The Hamming distance function $\textsc{Ham}_{n,d}$ returns 1 on all pairs of inputs x and y that differ in at most d coordinates and returns 0 otherwise. We initiate the study of the information complexity of the Hamming distance function.

We give a new optimal lower bound for the information complexity of the $\textsc{Ham}_{n,d}$ function in the small-error regime where the protocol is required to err with probability at most $\epsilon < d/n$. We also give a new conditional lower bound for the information complexity of $\textsc{Ham}_{n,d}$ that is optimal in all regimes. These results imply the first new lower bounds on the communication complexity of the Hamming distance function for the shared randomness two-way communication model since Pang and El-Gamal (1986). These results also imply new lower bounds in the areas of property testing and parity decision tree complexity.

1998 ACM Subject Classification F.1.2 Modes of Computation

Keywords and phrases Hamming distance, communication complexity, information complexity

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2014.465

1 Introduction

The Hamming distance function $\textsc{Ham}_{n,d} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ returns 1 on all pairs of inputs $x, y \in \{0,1\}^n$ that differ in at most d coordinates and returns 0 otherwise. This function is one of the fundamental objects of study in communication complexity. In this setting, Alice receives $x \in \{0,1\}^n$, Bob receives $y \in \{0,1\}^n$, and their goal is to compute the value of $\textsc{Ham}_{n,d}(x,y)$ while exchanging as few bits as possible.

The communication complexity of the Hamming distance function has been studied in various communication models [25, 18, 26, 11, 13], leading to tight bounds on the communication complexity of $\textsc{Ham}_{n,d}$ in many settings. One notable exception to this state of affairs is in the shared randomness two-way communication model in which Alice and Bob share a common source of randomness, they can both send messages to each other, and they are required to output the correct value of $\textsc{Ham}_{n,d}(x,y)$ with probability at least $1-\epsilon$ for each pair of inputs x, y. This can be done with a protocol that uses $O(\min\{n, d\log\frac{d}{\epsilon}\})$ bits of communication [13]. Furthermore, this protocol is quite simple: Alice and Bob simply take a random hash of their strings of length $O(\frac{d^2}{\epsilon})$ and determine if the Hamming distance of these hashes is at most d or not.

Pang and El-Gamal [18] showed that the hashing strategy is optimal when $d = cn$ for some constant $0 < c < 1$ and $0 < \epsilon < \frac{1}{2}$ is constant. With a simple padding argument, their result gives a general lower bound of $\Omega(\min\{d, n-d\})$ bits on the communication complexity of $\textsc{Ham}_{n,d}$.[1] Recently, there has been much interest in the Gap-Hamming Distance variant

[1] The same bound can also be obtained via a simple reduction from a promise version of the Set Disjointness

17th Int'l Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX'14) /
18th Int'l Workshop on Randomization and Computation (RANDOM'14).
Editors: Klaus Jansen, José Rolim, Nikhil Devanur, and Cristopher Moore; pp. 465–489

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

$\mathrm{GHD}_{n,d}$ of the Hamming distance function, where the inputs x and y are promised to be at Hamming distance at most $d - \sqrt{d}$ or at least $d + \sqrt{d}$ of each other. This line of work culminated in the recent proof that the $\Omega(\min\{d, n-d\})$ lower bound also holds for the $\mathrm{GHD}_{n,d}$ function [7, 22, 21]. Since Pang and El-Gamal's result, however, there has been no further progress on lower bounds for the communication complexity of the $\textsc{Ham}_{n,d}$ function and closing the gap between this lower bound and the upper bound of the simple hashing protocol remains an open problem.

In this work, we give new lower bounds on the communication complexity of the Hamming distance function by establishing new bounds on its information complexity. Informally, the information complexity of a function f is the amount of information that Alice and Bob must learn about each other's inputs when executing any protocol that computes f. The idea of using information complexity to lower bound the communication complexity of a function goes back to [8] and has since led to a number of exciting developments in communication complexity and beyond ([1, 2, 5, 24] to name just a few).

Let $\mathrm{IC}_\mu(f, \epsilon)$ denote the minimum amount of information that Alice and Bob can reveal to each other about their inputs while computing the function f with probability $1 - \epsilon$ (on every input pair), when their inputs are drawn from the distribution μ. The information complexity of f, denoted $\mathrm{IC}(f, \epsilon)$, is the maximum value of $\mathrm{IC}_\mu(f, \epsilon)$ over all distributions μ on the domain of f. A natural extension of the simple hashing protocol that gives the best-known upper bound on the communication complexity of $\textsc{Ham}_{n,d}$ also yields the best-known upper bound on its information complexity.

▶ **Proposition 1.1.** *For every* $0 < d < n - 1$ *and every* $0 \leq \epsilon < 1/2$,

$$\mathrm{IC}(\textsc{Ham}_{n,d}, \epsilon) \leq O(\min\{\log \tbinom{n}{d}, d \log \tfrac{d}{\epsilon}\}).$$

This bound on the information complexity of $\textsc{Ham}_{n,d}$ matches the communication complexity bound of the function when ϵ is a constant, but is exponentially smaller (in n) when d is small and ϵ tends to (or equals) 0.

By a reduction from a promise version of the Set Disjointness function and the known lower bound on the information complexity of that function [1], the information complexity of the Hamming distance problem is bounded below by

$$\mathrm{IC}(\textsc{Ham}_{n,d}, \epsilon) \geq \Omega(\min\{d, n-d\}) \tag{1}$$

for every $0 \leq \epsilon < \frac{1}{2}$. (In fact, Kerenidis et al. [15] have shown that the same lower bound also holds for the information complexity of the Gap-Hamming Distance function.) This result shows that the bound in Proposition 1.1 is optimal in the large distance regime, when $d = cn$ for some constant $0 < c < 1$.

The bound in Proposition 1.1 is also optimal when d and ϵ are both constants. In this case, the information complexity of $\textsc{Ham}_{n,d}$ is constant. There are two regimes, however, where the information complexity of the Hamming distance function is not yet well understood: the small-error regime where $\epsilon = o(1)$, and the medium-distance regime where $\omega(1) \leq d \leq o(n)$. In this paper, we introduce new lower bounds on the information complexity of $\textsc{Ham}_{n,d}$ for both of these regimes.

function. The optimal lower bound for the communication complexity of this function, however, was obtained later [14].

1.1 Our Results

1.1.1 Lower Bound for the Small-error Regime

Our first goal is to strengthen the lower bound on the information complexity of $\textsc{Ham}_{n,d}$ in the small-error regimes where $\epsilon = o(1)$ and where $\epsilon = 0$. It is reasonable to expect that for every value $0 \leq d \leq n-1$, the information complexity of every $\textsc{Ham}_{n,d}$ function should depend on either n or ϵ in these regimes. Surprisingly, Braverman [5] showed that this is not the case when $d = 0$. The $\textsc{Ham}_{n,0}$ function corresponds to the Equality function, and Braverman showed that for every $\epsilon \geq 0$, $\mathrm{IC}(\textsc{Equality}, \epsilon) = O(1)$ is bounded above by an absolute constant.

We show that the Equality function is in a sense a pathological example: it is the only Hamming distance function whose information complexity is independent of both n and ϵ.

▶ **Theorem 1.2.** *For every $1 \leq d < n-1$ and every $0 \leq \epsilon < 1/2$,*

$$\mathrm{IC}(\textsc{Ham}_{n,d}, \epsilon) = \Omega(\min\{\log \tbinom{n}{d}, d \log(1/\epsilon)\}).$$

The bound in the theorem matches that of Proposition 1.1 whenever $\epsilon < 1/n$. This shows that the lower bound is optimal in this regime and, notably, that the simple hashing protocol for $\textsc{Ham}_{n,d}$ is optimal among all protocols with low error.

There are two main components in the proof of Theorem 1.2. The first is a lower bound on the $\textsc{Ham}_{n,1\text{vs}.3}$, the promise version of the $\textsc{Ham}_{n,1}$ function where the protocol receives the additional guarantee that the two inputs x and y have Hamming distance exactly 1 or 3. Let μ be the uniform distribution over pairs (x, y) at Hamming distance 1 of each other. We show that every ϵ-error protocol for $\textsc{Ham}_{n,1\text{vs}.3}$ has large information cost over μ.

▶ **Lemma 1.3.** *Fix $\epsilon \geq 0$ and let μ be the uniform distribution over the pairs $(x, y) \sim \{0,1\}^n \times \{0,1\}^n$ at Hamming distance 1 of each other. Then*

$$\mathrm{IC}(\textsc{Ham}_{n,1vs.3}, \epsilon) \geq \mathrm{IC}_\mu(\textsc{Ham}_{n,1vs.3}, \epsilon) = \Omega(\min\{\log n, \log 1/\epsilon\}).$$

The second main component in the proof of Theorem 1.2 is a direct sum theorem (implicitly) due to Bar-Yossef et al. [1].[2] Roughly speaking, this direct sum theorem shows that under appropriate conditions, the information cost of any protocol that computes the AND of k copies of a function f is at least k times the information complexity of f. By observing that every protocol for the $\textsc{Ham}_{n,d}$ function also is a valid protocol for the AND of d copies of $\textsc{Ham}_{n/d,1\text{vs}.3}$, we are able to combine the direct sum theorem and Lemma 1.3 to complete the proof of Theorem 1.2.

1.1.2 Conditional Lower Bound

Theorem 1.2 establishes the optimality of the information complexity bound of Proposition 1.1 in every setting except the medium-distance regime, where $\omega(1) \leq d \leq o(n)$ and ϵ is (somewhat) large. We conjecture that the upper bound is optimal in this setting as well.

▶ **Conjecture 1.4.** *For every $1 \leq d < n-1$ and every $0 \leq \epsilon < 1/2$,*

$$\mathrm{IC}(\textsc{Ham}_{n,d}, \epsilon) = \Omega(\min\{\log \tbinom{n}{d}, d \log(d/\epsilon)\}).$$

[2] The direct sum theorem in [1] is stated for a different notion of information complexity but the proof of this theorem can be extended to yield a direct sum theorem for our setting as well. See Section 3 for the details.

A proof of the conjecture would have a number of interesting consequences. In particular, as we describe in more detail in Section 1.2.1 below, it would yield tight bounds on the communication complexity of $\textsc{Ham}_{n,d}$, on the query complexity of fundamental problems in property testing, and on the parity decision tree complexity of a natural Hamming weight function. A proof of the conjecture would also show that the simple hashing protocol is optimal and, in particular, since that protocol always accepts inputs at Hamming distance at most d from each other, it would confirm that two-sided error does not reduce the information or communication complexity of the Hamming distance function.

Finally, a proof of the conjecture would establish a notable separation between the communication complexity of Hamming distance and set disjointness. Let $\textsc{Disj}_n$ denote the function that returns 1 on the inputs $x, y \in \{0,1\}^n$ iff for every coordinate $i \in [n]$, $x_i = 0$ or $y_i = 0$. Let $\textsc{Disj}_{n,k}$ denote the variant on this problem where Alice and Bob's inputs are promised to have Hamming weight k. As mentioned briefly earlier, it is possible to get lower bounds on the communication complexity of $\textsc{Ham}_{n,d}$ with a reduction from $\textsc{Disj}_{n,(d+1)/2}$. When $d = cn$, and $0 < c < 1$ is a constant, this reduction is tight since both functions have communication complexity $\Theta(n)$ in this setting. However, Håstad and Wigderson [12] (see also [20]) showed that the communication complexity of $\textsc{Disj}_{n,k}$ is $O(k)$, so a proof of Conjecture 1.4 would show that the communication complexity of $\textsc{Ham}_{n,d}$ is asymptotically larger than that of $\textsc{Disj}_{n,(d+1)/2}$ when $d = o(n)$.

We give a conditional proof of Conjecture 1.4. To describe the result, we need to introduce a few notions related to parallel repetition. For a function $f : \{0,1\}^n \to \{0,1\}$ and $k \geq 2$, let $f^k : \{0,1\}^{nk} \to \{0,1\}^k$ denote the function that returns the value of f on k disjoint inputs. A protocol computes f^k with error ϵ if it computes the value of f on *all* k of the disjoint inputs with probability at least $1 - \epsilon$.

▶ **Definition 1.5.** A function $f : \mathcal{X}^n \times \mathcal{Y}^n \to \{0,1\}$ is *majority-hard* for the distribution μ on $\mathcal{X} \times \mathcal{Y}$ and for $\epsilon \geq 0$ if there exists a constant $c > 0$ such that for any $k \geq 2$,

$$\mathrm{IC}_{\mu^k}(\mathrm{Maj}_k \circ f, \epsilon) = \Theta(\mathrm{IC}_{\mu^{\lfloor ck \rfloor}}(f^{\lfloor ck \rfloor}, \epsilon)).$$

The upper bound in the definition trivially holds: a protocol for $\mathrm{Maj}_k \circ f$ can first determine the value of the k instances of f in parallel so $\mathrm{IC}_{\mu^k}(\mathrm{Maj}_k \circ f, \epsilon) \leq \mathrm{IC}_{\mu^k}(f^k, \epsilon)$. We believe that the reverse inequality holds for the $\textsc{Ham}_{n,1}$ function. In fact, we do not know of any distribution μ and any function f that is balanced on μ which is not majority-hard for μ. (Determining whether every such function is indeed majority-hard appears to be an interesting question in its own right; see [23] and [17] for related results.)

Let μ_1 and μ_3 be the uniform distributions over the pairs $(x, y) \in \{0,1\}^n \times \{0,1\}^n$ at Hamming distance 1 and 3 of each other, respectively. Let $\mu = \frac{1}{2}\mu_1 + \frac{1}{2}\mu_3$. We give a conditional proof of Conjecture 1.4 assuming that $\textsc{Ham}_{n,1}$ is a majority-hard function on μ.

▶ **Theorem 1.6.** *If* $\textsc{Ham}_{n,1}$ *is majority-hard over the distribution* μ *described above, then for every* $1 \leq d < n - 1$ *and every* $0 \leq \epsilon < 1/2$*,*

$$\mathrm{IC}(\textsc{Ham}_{n,d}, \epsilon) = \Theta(\min\{\log \tbinom{n}{d}, d \log(d/\epsilon)\}).$$

The proof of Theorem 1.6 follows the same overall structure as the proof of Theorem 1.2: we first establish a lower bound on the information complexity of $\textsc{Ham}_{n,1}$ and then use a direct sum theorem to derive the general lower bound from this result. Both of these components of the proof, however, must be significantly extended to yield the stronger lower bound.

In order to prove Theorem 1.6, we need to extend the result from Lemma 1.3 in two ways. First, we need extend the lower bound on the information complexity to apply to protocols in the average error model. In this model, a protocol has error ϵ under μ if the expected error probability on inputs drawn from μ. (By contrast, until now we have only considered protocols that must err with probability at most ϵ on every possible inputs; even those outside the support of μ.) Second, we need a lower bound that also applies to protocols that are allowed to abort with a constant probability δ. We denote the information complexity of the function f over the distribution μ in the ϵ-average-error δ-average-abortion-probability model by $\mathrm{IC}^{\mathrm{avg}}_{\mu}(f, \epsilon, \delta)$.

▶ **Lemma 1.7.** *Fix $0 \leq \epsilon < \frac{1}{2}$ and $0 \leq \delta < 1$. Let μ be the distribution described above. Then*

$$\mathrm{IC}^{\mathrm{avg}}_{\mu}(\textsc{Ham}_{n,1vs.3}, \epsilon, \delta) = \Omega(\min\{\log n, \log 1/\epsilon\}).$$

One significant aspect of the bound in Lemma 1.7 worth emphasizing is that the information complexity is *independent* of the abortion probability δ.

The second main component of the proof of Theorem 1.6 is another direct sum theorem. In this proof, we use a slightly different decomposition of $\textsc{Ham}_{n,d}$: instead of relating it to the composed function $\mathrm{AND}_d \circ \textsc{Ham}_{n/d,1\mathrm{vs}.3}$, we now use the fact that a protocol for $\textsc{Ham}_{n,d}$ also is a valid protocol for $\mathrm{Maj}_{d/2} \circ \textsc{Ham}_{2n/d,1\mathrm{vs}.3}$. If $\textsc{Ham}_{n,1}$ is majority-hard over the distribution μ, this decomposition shows that any protocol for $\textsc{Ham}_{n,d}$ has information complexity at least $\mathrm{IC}_{\mu^{d'}}(\textsc{Ham}^{d'}_{n,1\mathrm{vs}.3}, \epsilon, \delta)$ for some $d' = \Omega(d)$. We can then apply a recent strong direct sum theorem of Molinaro, Woodruff, and Yaroslavtsev [16] to obtain the desired result.

1.2 Extensions and Applications

1.2.1 Lower Bounds in Other Settings

The lower bounds on the information complexity of $\textsc{Ham}_{n,d}$ in Theorems 1.2 and 1.6 immediately imply corresponding lower bounds on the communication complexity of the same function.

▶ **Corollary 1.8.** *Fix $1 \leq d < n-1$ and $0 \leq \epsilon < \frac{1}{2}$. Then $R^{\mathrm{pub}}(\textsc{Ham}_{n,d}, \epsilon) = \Omega(\min\{\log \binom{n}{d}, d \log \frac{1}{\epsilon}\})$. Furthermore, if $\textsc{Ham}_{n,1}$ is majority-hard, then $R^{\mathrm{pub}}(\textsc{Ham}_{n,d}, \epsilon) = \Theta(\min\{\log \binom{n}{d}, d \log \frac{d}{\epsilon}\})$.*

In turn, the lower bounds on the communication complexity of $\textsc{Ham}_{n,d}$ imply new lower bounds on the query complexity of a number of different property testing problems via the connection introduced in [4].

▶ **Corollary 1.9.** *Fix $k \leq \frac{n}{2}$. At least $\Omega(\min\{k \log n, k \log \frac{1}{\delta}\})$ queries are required to test k-linearity and k-juntas with error δ. Furthermore, if $\textsc{Ham}_{n,1}$ is majority-hard, then $\Theta(k \log k)$ queries are required to test k-linearity and k-juntas with constant error.*

The best current lower bound on the query complexity for testing each property in Corollary 1.9 is $\Omega(k)$, a result that was obtained via a reduction from the Set Disjointness function [4]. Corollary 1.9 shows that replacing this reduction with one from the Hamming distance function yields stronger lower bounds.

Theorems 1.2 and 1.6 also give new lower bounds on the decision tree complexity of boolean functions. A *parity decision tree* is a tree where every internal node of the tree branches according to the parity of a specified subset of the bits of the input $x \in \{0,1\}^n$ and

every leaf is labelled with 0 or 1. The randomized ϵ-error parity decision tree complexity of a function $f : \{0,1\}^n \to \{0,1\}$, denoted $R^{\oplus}_{\epsilon}(f)$, is the minimum depth d such that there exists a distribution D over parity decision trees of depth d where for every $x \in \{0,1\}^n$, the path defined by x on a tree drawn from D leads to a leaf labelled by $f(x)$ with probability at least $1-\epsilon$. For $0 \leq d \leq n$, let $\textsc{Weight}_{n,d} : \{0,1\}^n \to \{0,1\}$ be the function that returns 1 iff the input x has Hamming weight at most d.

▶ **Corollary 1.10.** *Fix $0 < d < n-1$ and $0 \leq \epsilon < \frac{1}{2}$. Then $R^{\oplus}_{\epsilon}(\textsc{Weight}_{n,d}) = \Omega(\min\{\log \binom{n}{d}, d \log \frac{1}{\epsilon}\})$. Furthermore, if $\textsc{Ham}_{n,1}$ is majority-hard, then $R^{\oplus}_{\epsilon}(\textsc{Weight}_{n,d}) = \Theta(\min\{\log \binom{n}{d}, d \log \frac{d}{\epsilon}\})$.*

1.2.2 Symmetric XOR Functions

The Hamming distance functions $\textsc{Ham}_{n,d}$ are contained within a larger class of functions called *symmetric XOR functions*. The function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is a symmetric XOR function if it can be expressed as $f = h \circ \oplus_n$, where $\oplus_n : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^n$ is the entrywise XOR function and $h : \{0,1\}^n \to \{0,1\}$ is a symmetric boolean function.

The *skip complexity* of a symmetric XOR function $f = h \circ \oplus_n$ is defined as $\Gamma_{+2}(f) = \max\{0 \leq d < \frac{n}{2} : h(d) \neq h(d+2) \vee h(n-d) \neq h(n-d-2)\}$. This complexity measure is closely related to the Paturi complexity of symmetric functions [19]. The proof of Theorem 1.2 can be generalized to give a lower bound on the information complexity of every symmetric XOR function in terms of its skip complexity.

▶ **Theorem 1.11.** *Fix $\epsilon \geq 0$. For every symmetric XOR function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$,*

$$\mathrm{IC}(f, \epsilon) \geq \Omega(\Gamma_{+2}(f) \cdot \min\{\log n, \log 1/\epsilon\}).$$

The only symmetric XOR functions with skip complexity $\Gamma_{+2}(f) = 0$ are the affine combinations of the Equality and Parity functions. Each of these functions has information complexity $O(1)$, so Theorem 1.11 yields a complete characterization of the set of functions that have constant information complexity when $\epsilon = 0$.

1.2.3 Direct Sum Violations

In 1995, Feder et al. [10] showed that the Equality function violates the direct-sum theorem in the randomized communication complexity model when $\epsilon = o(1)$. Braverman [5] noted that an alternative proof of this fact follows from the fact that the information complexity of the Equality function satisfies $\mathrm{IC}(\textsc{Equality}, \epsilon) = O(1)$.

The tight characterization of the information complexity of $\textsc{Ham}_{n,1}$ obtained by the bounds in Proposition 1.1 and Lemma 1.3 shows that $\textsc{Ham}_{n,1}$ satisfies the direct-sum theorem for randomized communication complexity when $n = \mathrm{poly}(1/\epsilon)$ and violates it otherwise (i.e., when $\log n = o(\log 1/\epsilon)$. This result can be seen as further evidence of the qualitative difference between the complexity of the Equality function and that of the "almost-equality" function $\textsc{Ham}_{n,1}$. See Section 7 for the details.

1.2.4 Composition of the $\textsc{Ham}_{n,1}$ Function

One important difference between the proof of Theorem 1.2 and that of Theorem 1.6 is that whereas the former is obtained by analyzing the composed function $\mathrm{AND}_d \circ \textsc{Ham}_{n,1\mathrm{vs}.3}$, the latter is obtained by analyzing $\mathrm{Maj}_{d/2} \circ \textsc{Ham}_{n,1\mathrm{vs}.3}$. It is natural to ask whether this

switch is necessary—whether the stronger lower bound of Theorem 1.6 could be obtained by considering the composed function $\mathrm{AND}_d \circ \mathrm{HAM}_{n,1\mathrm{vs}.3}$.

The same question can be rephrased to ask whether the bound in Theorem 1.2 is optimal for the function $\mathrm{AND}_d \circ \mathrm{HAM}_{n,1\mathrm{vs}.3}$. We show that it is. Furthermore, we show that a similar upper bound also applies to the function $\mathrm{OR}_k \circ \mathrm{HAM}_{n,1}$, so that in order to obtain the lower bound in Theorem 1.6 via a reduction approach, we must consider another composition function. See Section 8 for the details.

2 Information Complexity Preliminaries

We use standard information-theoretic notation and the following basic facts about entropy and mutual information. See [9] for the basic definitions and the proofs of the following facts.

▶ **Fact 2.1.** *If X can be described with k bits given Y, then $H(X|Y) \leq k$.*

▶ **Fact 2.2.** $I(X,Y|Z) = H(X|Z) - H(X|Y,Z)$.

▶ **Fact 2.3** (Chain rule for conditional mutual information)**.** $I(X_1,X_2;Y|Z) = I(X_1;Y|Z) + I(X_2;Y|X_1,Z)$.

▶ **Fact 2.4** (Data processing inequality)**.** *If $I(X;Z|Y,W) = 0$, then $I(X;Y|W) \geq I(X;Z|W)$.*

▶ **Fact 2.5.** *If $I(X;W|Y,Z) = 0$, then $I(X;Y|Z) \geq I(X;Y|Z,W)$.*

▶ **Definition 2.6** (Kullback–Leibler divergence)**.** The *Kullback–Leibler (KL) divergence* between two distributions μ, ν is $D_{\mathrm{KL}}(\mu \parallel \nu) = \sum_x \mu(x) \log \frac{\mu(x)}{\nu(x)}$.

▶ **Fact 2.7** (Gibbs' inequality)**.** *For every distributions μ and ν, $D_{\mathrm{KL}}(\mu \parallel \nu) \geq 0$.*

▶ **Fact 2.8.** *For any distribution μ on $\mathcal{X} \times \mathcal{Y}$ with marginals μ_X and μ_Y, the mutual information of the random variables $(A,B) \sim \mu$ satisfies $I(A;B) = D(\mu \parallel \mu_X \mu_Y)$.*

▶ **Fact 2.9** (Log-sum inequality)**.** *Let $n \in \mathbb{N}$ and $a_1,\ldots,a_n,b_1,\ldots,b_n$ be non-negative real numbers. Define $A := \sum_{i=1}^n a_i$ and $B := \sum_{i=1}^n b_i$. Then, $\sum_{i=1}^n a_i \log(a_i/b_i) \geq A \log(A/B)$.*

▶ **Definition 2.10** (Information cost)**.** Let μ be a distribution with support $\{0,1\}^n \times \{0,1\}^n$ and let $(X,Y) \sim \mu$ where X is Alice's input and Y is Bob's input. The *information cost* of a protocol Π with respect to μ is defined by $\mathrm{IC}_\mu(\Pi) := I_\mu(\Pi(X,Y);X|Y) + I_\mu(\Pi(X,Y);Y|X)$.

▶ **Definition 2.11** (Prior-free information complexity)**.** Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be a function and let $\epsilon > 0$. The *prior-free information complexity* of f with error rate ϵ is defined by $\mathrm{IC}(f,\epsilon) := \min_\Pi \max_\mu \mathrm{IC}_\mu(\Pi)$ where Π ranges over all protocols computing f with error probability at most ϵ on each input pair in $\{0,1\}^n \times \{0,1\}^n$ and μ ranges over all distributions with support $\{0,1\}^n \times \{0,1\}^n$.

▶ Remark. Braverman [5] distinguished between internal information measures that quantify the amount of information that Alice and Bob reveal to each other and external information measures that quantify the amount of information that Alice and Bob reveal to an external observer. Definitions 2.10 and 2.11 refer to the *internal* information cost and *internal* prior-free information complexity respectively.

3 Lower Bound for the Small-error Regime

In this section, we complete the proof of Theorem 1.2, giving an unconditional lower bound on the information complexity of $\textsc{Ham}_{n,d}$. In fact, we do more: we show that the same information complexity lower bound holds even for protocols that receive the additional promise that every block of n/d coordinates in $[n]$ contains exactly 1 or 3 coordinates on which x and y differ. Furthermore, we show that our information complexity lower bound holds under the distribution where we choose the inputs x and y uniformly at random from all such pairs of inputs that have Hamming distance exactly 1 on each block.

The proof has two main components. The first is our lower bound on the information complexity of the $\textsc{Ham}_{n,1\text{vs}.3}$ function, which is the more technically challenging component of the proof and which we defer to the next subsection. The second is a direct sum theorem for information complexity. In order to state this theorem, we must first introduce a bit more notation. We use $[n]$ to denote the set $\{1,\dots,n\}$. For $X = X_1X_2\cdots X_n \in \mathcal{X}^n$ and $i < k < n$, let $X_{[k]}$ and $X_{[i:k]}$ denote the strings $X_1\cdots X_k$ and $X_i\cdots X_k$ respectively. For $i \in [n]$, we use e_i to denote the n-bit string $z \in \{0,1\}^n$ with $z_i = 1$ and all other bits $z_j = 0$.

▶ **Definition 3.1** (Composed function)**.** The *composition* of the functions $f : \{0,1\}^k \to \{0,1\}$ and $g : \mathcal{X}\times\mathcal{Y} \to \{0,1\}$ is the function $f\circ g : \mathcal{X}^k\times\mathcal{Y}^k \to \{0,1\}$ defined by $(f\circ g)(x,y) = f\big(g(x_1,y_1),\dots,g(x_k,y_k)\big)$.

▶ **Definition 3.2.** For a vector $x \in \mathcal{X}^k$, an index $j \in [k]$, and an element $u \in \mathcal{X}$, define $x_{j\leftarrow u}$ to be the vector in $\mathcal{X}^k$ obtained by replacing the jth coordinate of x with u.

▶ **Definition 3.3** (Collapsing distributions)**.** A distribution μ over $\mathcal{X}\times\mathcal{Y}$ is a *collapsing distribution* for the composed function $f\circ g : \mathcal{X}^k\times\mathcal{Y}^k \to \{0,1\}$ if every point (x,y) in the support of μ, every $j \in [k]$, and every $(u,v) \in \mathcal{X}\times\mathcal{Y}$ satisfy $f\circ g(x_{j\leftarrow u}, y_{j\leftarrow v}) = g(u,v)$.

We use the following direct-sum theorem, which is essentially due to Bar-Yossef et al. [1] and to Braverman and Rao [6]. We include the proof for the convenience of the reader.

▶ **Theorem 3.4** (Direct-sum theorem)**.** *Let μ^k be a collapsing distribution for the composed function $f\circ g : \mathcal{X}^k\times\mathcal{Y}^k \to \{0,1\}$. For every $\epsilon \geq 0$, $\mathrm{IC}_{\mu^k}(f\circ g,\epsilon) \geq k\,\mathrm{IC}_\mu(g,\epsilon)$.*

Proof. Consider an ϵ-error protocol P for $f\circ g$ with optimal information cost over μ^k. Let $\Pi(x,y)$ be a random variable (over the private randomness of the protocol) denoting the transcript of the protocol on inputs $x,y \in \mathcal{X}^k\times\mathcal{Y}^k$. By the optimality of P and two applications of the chain rule for mutual information in opposite directions,

$$\begin{aligned}\mathrm{IC}_{\mu^k}(f\circ g,\epsilon) &= I(X;\Pi(X,Y)\mid Y) + I(Y;\Pi(X,Y)\mid X)\\ &= \sum_{i=1}^{k} I(X_i;\Pi(X,Y)\mid Y, X_{[i-1]}) + I(Y_i;\Pi(X,Y)\mid X, Y_{[i+1,k]}).\end{aligned}$$

Since $I(X_i;Y_{[i-1]}\mid X_{[i-1]},Y_{[i,k]}) = 0$, we have $I(X_i;\Pi(X,Y)\mid Y,X_{[i-1]}) \geq I(X_i;\Pi(X,Y)\mid X_{[i-1]},Y_{[i,k]})$. Similarly, $I(Y_i;\Pi(X,Y)\mid X,Y_{[i+1,k]}) \geq I(Y_i;\Pi(X,Y)\mid X_{[i]},Y_{[i+1,k]})$. So

$$\mathrm{IC}_{\mu^k}(f\circ g,\epsilon) \geq \sum_{i=1}^{k} I(X_i;\Pi(X,Y)\mid X_{[i-1]}Y_{[i,k]}) + I(Y_i;\Pi(X,Y)\mid X_{[i]}Y_{[i+1,k]}).$$

To complete the proof, we want to show that each summand is the information cost of an ϵ-error protocol for g over μ. Fix an index $i \in [k]$. Let P_i^* be a protocol that uses the

public randomness to draw $X'_1, \ldots, X'_{i-1}$ from the marginal of μ on $\mathcal{X}$ and $Y'_{i+1}, \ldots, Y'_k$ from the marginal of μ on $\mathcal{Y}$. Alice draws $X'_{i+1}, \ldots, X'_k$ using her private randomness so that $(X'_{i+1}, Y'_{i+1}), \ldots, (X'_k, Y'_k) \sim \mu$. Similarly, Bob uses his private randomness to draw $Y'_1, \ldots, Y'_{i-1}$ such that $(X'_1, Y'_1), \ldots, (X'_{i-1}, Y'_{i-1}) \sim \mu$. They then set $X'_i \leftarrow X_i$ and $Y'_i \leftarrow Y_i$. The protocol P_i^* then simulates P on (X', Y') and returns the value of $f \circ g(X', Y')$. Since μ^k is a collapsing distribution, $g(X_i, Y_i) = f \circ g(X', Y')$ and P_i^* is a valid ϵ-error protocol for g. In turn, this implies that

$$\begin{aligned}\mathrm{IC}_{\mu^k}(f \circ g, \epsilon) &\geq \sum_{i=1}^{k} I(X_i; \Pi(X,Y) \mid X_{[i-1]} Y_{[i,k]}) + I(Y_i; \Pi(X,Y) \mid X_{[i]} Y_{[i+1,k]}) \\ &\geq \sum_{i=1}^{k} \mathrm{IC}_{\mu}(g, \epsilon) = k \, \mathrm{IC}_{\mu}(g, \epsilon).\end{aligned}$$

◀

Let μ be the uniform distribution on pairs $(x, y) \in \{0,1\}^n \times \{0,1\}^n$ at Hamming distance one from each other. In the following subsection, we show that every protocol for $\textsc{Ham}_{n,1\text{vs}.3}$ must have information complexity $\Omega(\min\{\log n, \log \frac{1}{\epsilon}\})$ under this distribution. We can then apply the direct sum theorem to complete the proof of Theorem 1.2.

Proof of Theorem 1.2. Any protocol for $\textsc{Ham}_{n,d}$ also is a valid protocol for the composed function $\mathrm{AND}_d \circ \textsc{Ham}_{n/d,1\text{vs}.3}$. So for every $\epsilon \geq 0$,

$$\mathrm{IC}(\textsc{Ham}_{n,d}, \epsilon) \geq \mathrm{IC}(\mathrm{AND}_d \circ \textsc{Ham}_{n/d,1\text{vs}.3}, \epsilon).$$

Let μ be the uniform distribution on pairs $(x, y) \in \{0,1\}^{n/d} \times \{0,1\}^{n/d}$ with Hamming distance 1. By definition, $\mathrm{IC}(\mathrm{AND}_d \circ \textsc{Ham}_{n/d,1\text{vs}.3}, \epsilon) \geq \mathrm{IC}_{\mu^d}(\mathrm{AND}_d \circ \textsc{Ham}_{n/d,1\text{vs}.3}, \epsilon)$. Moreover, since the support of μ is on pairs x, y at Hamming distance 1 from each other, μ^d is a collapsing distribution for $AND_d \circ \textsc{Ham}_{n/d,1\text{vs}.3}$. So by Theorem 3.4,

$$\mathrm{IC}_{\mu^d}(\mathrm{AND}_d \circ \textsc{Ham}_{n/d,1\text{vs}.3}, \epsilon) \geq d \, \mathrm{IC}_{\mu}(\textsc{Ham}_{n/d,1\text{vs}.3}, \epsilon)$$

and the theorem follows from Lemma 1.3. ◀

3.1 Proof of Lemma 1.3

In this section, we give a lower bound on the information complexity of protocols for $\textsc{Ham}_{n,1\text{vs}.3}$ under the distribution μ that is uniform over the pairs of vectors $(x, y) \in \{0,1\}^n \times \{0,1\}^n$ at Hamming distance 1 from each other.

▶ **Fact 3.5** (Rectangle bound [1]). *For any protocol whose transcript on inputs x, y (resp., x', y') is the random variable $\Pi(x, y)$ (resp., $\Pi(x', y')$) and for any possible transcript t,*

$$\Pr[\Pi(x,y) = t] \Pr[\Pi(x',y') = t] = \Pr[\Pi(x,y') = t] \Pr[\Pi(x',y) = t].$$

▶ **Fact 3.6** (Extension of Gibbs' inequality). *For every distributions μ and ν on $\mathcal{X}$, and every subset $S \subseteq \mathcal{X}$, $\sum_{x \in S} \mu(x) \log \frac{\mu(x)}{\nu(x)} \geq \ln 2 \, (\mu(S) - \nu(S))$.*

Proof. Using the inequality $\log x \leq \ln 2 (x - 1)$, we obtain

$$\sum_{x \in S} \mu(x) \log \frac{\mu(x)}{\nu(x)} = -\sum_{x \in S} \mu(x) \log \frac{\nu(x)}{\mu(x)} \geq \sum_{x \in S} \mu(x) \ln 2 \, (1 - \frac{\nu(x)}{\mu(x)}) \geq \ln 2 \, (\mu(S) - \nu(S)).$$ ◀

▶ **Lemma 3.7.** *Let Π be a randomized protocol and let T be the set of all possible transcripts of Π. Let μ be the uniform distribution on pairs $(x,y) \in \{0,1\}^n \times \{0,1\}^n$ at Hamming distance 1 from each other. Then*

$$IC_\mu(\Pi(X,Y)) = \mathop{\mathbb{E}}_{z\in\{0,1\}^n, i\in[n]} \sum_{t\in T} \Pr[\Pi(z\oplus e_i, z) = t] \log \frac{\Pr[\Pi(z\oplus e_i, z) = t]}{\mathbb{E}_{j,\ell\in[n]} \Pr[\Pi(z\oplus e_i\oplus e_j, z\oplus e_\ell) = t]}.$$

Proof. The mutual information of X and $\Pi(X,Y)$ given Y satisfies

$$\begin{aligned}
&I(X;\Pi(X,Y) \mid Y)\\
&\quad= \mathop{\mathbb{E}}_{y}[I(X;\Pi(X,y) \mid Y = y)]\\
&\quad= \mathop{\mathbb{E}}_{y}[D_{\mathrm{KL}}(X,\Pi(X,y) \parallel X,\Pi(X',y))]\\
&\quad= \mathop{\mathbb{E}}_{y}\left[\sum_{x\in\{0,1\}^n}\sum_{t\in T} \Pr[X=x]\Pr[\Pi(x,y)=t]\log\frac{\Pr[X=x]\Pr[\Pi(x,y)=t]}{\Pr[X=x]\Pr[\Pi[X',y]=t]}\right]\\
&\quad= \mathop{\mathbb{E}}_{z,i}\left[\sum_{t\in T}\Pr[\Pi(z\oplus e_i,z)=t]\log\frac{\Pr[\Pi(z\oplus e_i,z)=t]}{\mathbb{E}_{\ell\in[n]}\Pr[\Pi(z\oplus e_\ell,z)=t]}\right].
\end{aligned}$$

Similarly,

$$I(Y;\Pi(X,Y) \mid X) = \mathop{\mathbb{E}}_{z,i}\left[\sum_{t\in T}\Pr[\Pi(z\oplus e_i,z)=t]\log\frac{\Pr[\Pi(z\oplus e_i,z)=t]}{\mathbb{E}_{\ell\in[n]}\Pr[\Pi(z\oplus e_i,z\oplus e_i\oplus e_j)=t]}\right].$$

Summing those two expressions, we obtain

$$IC_\mu(\Pi(X,Y)) = \mathop{\mathbb{E}}_{z,i}\left[\sum_{t\in T}\Pr[\Pi(z\oplus e_i,z)=t] \log\frac{\Pr[\Pi(z\oplus e_i,z)=t]^2}{\mathbb{E}_{j,\ell\in[n]}\Pr[\Pi(z\oplus e_\ell,z)=t]\Pr[\Pi(z\oplus e_i,z\oplus e_i\oplus e_j)=t]}\right].$$

By the rectangle bound (Fact 3.5),

$$\Pr[\Pi(z\oplus e_\ell,z)=t]\Pr[\Pi(z\oplus e_i,z\oplus e_i\oplus e_j)=t] = \Pr[\Pi(z\oplus e_i,z)=t]\Pr[\Pi(z\oplus e_\ell,z\oplus e_i\oplus e_j)=t]$$

and the lemma follows. ◀

Proof of Lemma 1.3. Fix any ϵ-error protocol for $\textsc{Ham}_{n,1\text{vs}.3}$. Let $\Pi(x,y)$ denote (a random variable representing) its transcript on inputs x,y. Let T^1 denote the set of transcripts for which the protocol outputs 1. By Lemma 3.7 and the extended Gibbs' inequality (Fact 3.6),

$$\mathrm{IC}_\mu(\Pi(X,Y)) \geq \mathop{\mathbb{E}}_{z\in\{0,1\}^n, i\in[n]} \sum_{t\in T^1} \Pr[\Pi(z\oplus e_i,z)=t] \log\frac{\Pr[\Pi(z\oplus e_i,z)=t]}{\mathbb{E}_{j,\ell\in[n]}\Pr[\Pi(z\oplus e_i\oplus e_j,z\oplus e_\ell)=t]} - \ln 2\,.$$

The correctness of the protocol guarantees that when i,j,ℓ are all disjoint, then $\sum_{t\in T^1}\Pr[\Pi(z\oplus e_i\oplus e_j, z\oplus e_\ell)=t] \leq \epsilon$. For any $z\in\{0,1\}^n$ and $i\in[n]$, the probability that i,j,ℓ are all disjoint is $(n-1)(n-2)/n^2 > 1-3/n$. Therefore,

$$\sum_{t\in T^1}\mathop{\mathbb{E}}_{j,\ell\in[n]}\Pr[\Pi(z\oplus e_i\oplus e_j, z\oplus e_\ell)=t] \leq 3/n+\epsilon$$

and by the log-sum inequality and the fact that $x \log_2(x) \geq -0.6$ for all $x \in [0,1]$,

$$\begin{aligned}\mathrm{IC}_\mu(\Pi(X,Y)) &\geq \Pr[\Pi(z \oplus e_i, z) \in T^1] \log \frac{\Pr[\Pi(z \oplus e_i, z) \in T^1]}{\mathbb{E}_{j,\ell} \Pr[\Pi(z \oplus e_i \oplus e_j, z \oplus e_\ell) \in T^1]} \\ &\geq (1-\epsilon) \log \frac{1-\epsilon}{3/n+\epsilon} - \ln 2 \geq (1-\epsilon) \log \frac{1}{3/n+\epsilon} - O(1)\,. \end{aligned}$$

◀

4 Conditional Lower Bound

In this section, we prove Theorem 1.6. We will need the following notion of information complexity.

▶ **Definition 4.1** (Information complexity with average-case abortion and error)**.** Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$. Then, $\mathrm{IC}_{\mu,\delta,\epsilon}(f|\nu)$ is the minimum conditional information cost of a randomized protocol that computes f with abortion probability at most δ and error probability at most ϵ, where the probabilities are taken over both the internal (public and private) randomness of the protocol Π and over the randomness of the distribution μ.

We now give the slight generalization of the MWY theorem that we will use to prove Theorem 1.6.

▶ **Theorem 4.2** (Slight generalization of the direct-sum theorem of [16])**.** *Let $X \in \mathcal{X}$, $Y \in \mathcal{Y}$ and λ be a distribution on (X,Y,D) with marginals μ over (X,Y) and ν over D such that for every value d of D, X and Y are conditionally independent given $D = d$. For any $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, $k \in \mathbb{N}$ and $\epsilon \leq 1/3$, $\mathrm{IC}_{\mu^k,\epsilon}(f^k|\nu^k) = k \cdot \Omega(\mathrm{IC}_{\mu,O(\epsilon),O(\epsilon/k)}(f|\nu))$.*

Proof. See appendix A for the proof and the comparison to the direct-sum theorem of [16]. ◀

We will lower bound the information revealed by any protocol computing $\mathrm{HAM}_{n,1}$ with small error and abortion with respect to some hard input distribution. Here, the error and abortion probabilities are over both the hard input distribution and the public and private randomness of the protocol. We handle abortion probabilities and allow such average-case guarantees in order to be able to apply Theorem 4.2. We first define our hard input probability distribution. We define the distribution λ over tuples (B,D,Z,I,J,L,X,Y) as follows: To sample $(B,D,Z,I,J,L,X,Y) \sim \lambda$, we sample $B, D \in_R \{0,1\}$, $Z \in_R \{0,1\}^n$, $I,J,L \in_R [n]$ and:

- If $B = 0$,
  - If $D = 0$, set $(X,Y) = (Z, Z \oplus e_I)$.
  - If $D = 1$, set $(X,Y) = (Z \oplus e_I, Z)$.
- If $B = 1$,
  - If $D = 0$, set $(X,Y) = (Z \oplus e_I \oplus e_J, Z \oplus e_L)$.
  - If $D = 1$, set $(X,Y) = (Z \oplus e_L, Z \oplus e_I \oplus e_J)$.

We let μ be the marginal of λ over (X,Y) (and ν be the marginal of λ over (B,D,Z)). Note that conditioned on B, D and Z taking any particular values, X and Y are independent. That is, we have a mixture of product distributions. We will prove the following lemma (which is a stronger version of Lemma 1.7).

▶ **Lemma 4.3.** *Let Π be a randomized protocol that computes $\mathrm{HAM}_{n,1}$ with abortion probability at most δ and error probability at most ϵ, where the probabilities are taken over both the internal (public and private) randomness of the protocol Π and over the randomness of our*

marginal distribution μ. Let q and w be such that $4/q + 4(\delta+\epsilon)/w \leq 1$ and $w \leq 1$. Then, we have that

$$I((X,Y);\Pi(X,Y)|Z,D,B=0) \geq (1-\frac{4}{q}-\frac{4(\delta+\epsilon)}{w})\frac{(1-w)}{2}\log_2(\frac{1}{3/n+q\epsilon}) - O(1)\,. \quad (2)$$

For $\delta \leq 1/32$ and $\epsilon \leq 1/32$, setting $w = 16(\delta+\epsilon)$ and $q = 16$ in inequality (2) yields

$$\begin{aligned} I((X,Y);\Pi(X,Y)|Z,D,B) &= \Omega(I((X,Y);\Pi(X,Y)|Z,D,B=0)) \\ &= \Omega(\min(\log n, \log(1/\epsilon))) - O(1)\,. \end{aligned}$$

Given Lemma 4.3, we can now complete the proof of Theorem 1.6.

Proof of Theorem 1.6. Since $\textsc{Ham}_{n,d} = \textsc{Ham}_{n,n-d}$, it suffices to prove the bound for $d \leq n/2$. Applying Theorem 4.2 with $f = \textsc{Ham}_{n/d,1}$, $k = d$ and the distributions μ and ν given above, we get that

$$\mathrm{IC}_{\mu^d,\epsilon}((\textsc{Ham}_{n/d,1})^d|\nu^d) = d \cdot \Omega(\mathrm{IC}_{\mu,O(\epsilon),O(\epsilon/d)}(\textsc{Ham}_{n/d,1}|\nu)).$$

By Lemma 4.3, we also have that

$$\mathrm{IC}_{\mu,O(\epsilon),O(\epsilon/d)}(\textsc{Ham}_{n/d,1}|\nu) = \Omega(\min(\log(n/d), \log(d/\epsilon))) - O(1).$$

Hence,

$$\mathrm{IC}_{\mu^d,\epsilon}((\textsc{Ham}_{n/d,1})^d|\nu^d) = d \cdot \Omega(\min(\log(n/d), \log(d/\epsilon))) - O(d).$$

Using the assumption that $\textsc{Ham}_{n/d,1}$ is majority-hard, Theorem 1.6 now follows. ◀

Given Lemma 4.3, we can also complete the proof of Lemma 1.7.

Proof of Lemma 1.7. Let Π be a randomized protocol that computes $\textsc{Ham}_{n,1}$ with abortion probability at most δ and error probability at most ϵ, where the probabilities are taken over both the internal (public and private) randomness of the protocol Π and over the randomness of our marginal distribution μ. We have that

$$\begin{aligned} \mathrm{IC}_\mu(\Pi) &= I_\mu(\Pi(X,Y);X|Y) + I_\mu(\Pi(X,Y);Y|X) \\ &\overset{(a)}{\geq} I_\lambda(\Pi(X,Y);X|Y,D,B) + I_\lambda(\Pi(X,Y);Y|X,D,B) \\ &\geq \frac{1}{4}(I_\lambda(\Pi(X,Y);X|Y,D=1,B=0) + I_\lambda(\Pi(X,Y);Y|X,D=0,B=0)) \\ &= \frac{1}{4}(I_\lambda(\Pi(X,Y);X|Z,D=1,B=0) + I_\lambda(\Pi(X,Y);Y|Z,D=0,B=0)) \\ &= \frac{1}{2}I_\lambda(\Pi(X,Y);X|Z,D,B=0) \\ &\overset{(b)}{=} \Omega(\min(\log n, \log(1/\epsilon))) - O(1). \end{aligned}$$

where (a) follows from Fact 2.5 and the fact that $I(\Pi(X,Y);(D,B)|X,Y) = 0$ and (b) follows from Lemma 4.3. ◀

4.1 Proof of Lemma 4.3

We start by sketching the idea of the proof of Lemma 4.3 before giving the full proof. We first note that the conditional information cost that we want to lower bound can be expressed as an average, over a part of the input distribution, of a quantity that still carries the randomness of the protocol. We show that most distance-1 input pairs are computed correctly and have an expected error probability over their distance-3 "cousin pairs"[3] of at most $O(\epsilon)$. We can thus average over only such distance-1 input pairs at the cost of a multiplicative constant-factor decrease in the lower bound. At this point, the remaining randomness is due solely to the protocol. It turns out that we can deal with the corresponding quantity in a similar way to how we dealt with the randomness in the proof of Lemma 1.3, i.e., using the extended Gibbs' inequality and the log-sum inequality. We now give the full proof.

Proof of Lemma 4.3. Let T be the set of all possible transcripts of Π. By Lemma 3.7, we have that[4]

$$\begin{aligned} &I((X,Y);\Pi|Z,D,B=0) \\ &\quad= \frac{1}{2}\mathop{\mathbb{E}}_{z\in\{0,1\}^n,i\in[n]}\sum_{t\in T}\Pr[\Pi(z\oplus e_i,z)=t]\log\frac{\Pr[\Pi(z\oplus e_i,z)=t]}{\mathbb{E}_{j,\ell\in[n]}\Pr[\Pi(z\oplus e_i\oplus e_j,z\oplus e_\ell)=t]} \\ &\quad= \frac{1}{2}\mathop{\mathbb{E}}_{z\in\{0,1\}^n,i\in[n]}\kappa_{z,i} \end{aligned}$$

with

$$\kappa_{z,i} := \sum_{t\in T}\Pr[\Pi(z\oplus e_i,z)=t]\log\frac{\Pr[\Pi(z\oplus e_i,z)=t]}{\mathbb{E}_{j,\ell\in[n]}\Pr[\Pi(z\oplus e_i\oplus e_j,z\oplus e_\ell)=t]}.$$

By the log-sum inequality, we have:

▶ **Fact 4.4.** *For every* $(z,i)\in\{0,1\}^n\times[n]$, $\kappa_{z,i}\geq 0$.

Let q and w be such that $4/q+4(\delta+\epsilon)/w\leq 1$ and $w\leq 1$.

▶ **Definition 4.5** (Nice (z,i)-pairs)**.** A pair $(z,i)\in\{0,1\}^n\times[n]$ is said to be *nice* if it satisfies the following two conditions:

1. $\Pr_{\Pi,j,l\in[n]}[\Pi(z\oplus e_i\oplus e_j,z\oplus e_\ell)\neq \textsc{Ham}_{n,1}(z\oplus e_i\oplus e_j,z\oplus e_\ell)$ and $\Pi(z\oplus e_i\oplus e_j,z\oplus e_\ell)$ does not abort] is at most $q\epsilon$.
2. $\Pr_\Pi[\Pi(z\oplus e_i,z)]\neq \textsc{Ham}_{n,1}(z\oplus e_i,z)]\leq w$

The following lemma shows that most (z,i)-pairs are nice:

▶ **Lemma 4.6.** *The fraction of pairs* $(z,i)\in\{0,1\}^n\times[n]$ *that are nice is at least* $1-4/q-4(\delta+\epsilon)/w$.

[3] For a distance-1 input pair $(z\oplus e_i,z)$, its distance-3 "cousin pairs" are those of the form $(z\oplus e_i\oplus e_j,z\oplus e_\ell)$ for $j,\ell\in[n]$. Note that this step uses the two-sided nature of our new distribution.

[4] Note that given $B=0$, (X,Y) is a uniformly-random distance-1 pair. Thus, $I((X,Y);\Pi(X,Y)|Z,D,B=0)$ is equal to the internal information complexity $\mathrm{IC}_\mu(\Pi(X,Y))$ in Lemma 3.7 up to a multiplicative factor of 2.

Proof of Lemma 4.6. We have that

$$\begin{aligned}
&\mathop{\mathbb{E}}_{z,i}[\Pr_{\Pi,j,l}[\Pi(z\oplus e_i\oplus e_j, z\oplus e_\ell)\neq \textsc{Ham}_{n,1}(z\oplus e_i\oplus e_j, z\oplus e_\ell)\\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{and } \Pi(z\oplus e_i\oplus e_j, z\oplus e_\ell) \text{ does not abort}]]\\
=\ &\Pr_{z,i,\Pi,j,l}[\Pi(z\oplus e_i\oplus e_j, z\oplus e_\ell)\neq \textsc{Ham}_{n,1}(z\oplus e_i\oplus e_j, z\oplus e_\ell)\\
&qquad\qquad\qquad\qquad\qquad\qquad \text{and } \Pi(z\oplus e_i\oplus e_j, z\oplus e_\ell) \text{ does not abort}]\\
\leq\ &4\Pr_{\Pi,(x,y)\sim\mu}[\Pi(x,y)\neq \textsc{Ham}_{n,1}(x,y) \text{ and } \Pi(x,y) \text{ does not abort}]\\
\leq\ &4\epsilon.
\end{aligned}$$

Thus, by Markov's inequality, the fraction of (z,i)-pairs for which

$$\begin{aligned}
\Pr_{\Pi,j,l}[&\Pi(z\oplus e_i\oplus e_j, z\oplus e_\ell)\neq \textsc{Ham}_{n,1}(z\oplus e_i\oplus e_j, z\oplus e_\ell)\\
&\qquad\qquad \text{and } \Pi(z\oplus e_i\oplus e_j, z\oplus e_\ell) \text{ does not abort}] > q\epsilon
\end{aligned}$$

is at most $4/q$. Moreover, we have that

$$\begin{aligned}
\mathop{\mathbb{E}}_{z,i}[\Pr_{\Pi}[\Pi(z\oplus e_i, z)\neq \textsc{Ham}_{n,1}(z\oplus e_i, z)]] &= \Pr_{\Pi,z,i}[\Pi(z\oplus e_i, z)\neq \textsc{Ham}_{n,1}(z\oplus e_i, z)]\\
&\leq 4\Pr_{\Pi,(x,y)\sim\mu}[\Pi(x,y)\neq \textsc{Ham}_{n,1}(x,y)]\\
&\leq 4(\delta+\epsilon).
\end{aligned}$$

Applying Markov's inequality once again, we get that the fraction of (z,i)-pairs for which

$$\Pr_{\Pi}[\Pi(z\oplus e_i, z)\neq \textsc{Ham}_{n,1}(z\oplus e_i, z)]\geq w$$

is at most $4(\delta+\epsilon)/w$. By the union bound, we conclude that the fraction of (z,i)-pairs that are nice is at least $1-4/q-4(\delta+\epsilon)/w$. ◀

Let $N\subseteq\{0,1\}^n\times[n]$ be the set of all nice (z,i)-pairs. Using the fact that $\kappa_{z,i}\geq 0$ for all z and i (Fact 4.4), we get that:

$$I((X,Y);\Pi(X,Y)|Z,D,B=0)\geq\frac{1}{2}\frac{|N|}{n2^n}\mathop{\mathbb{E}}_{(z,i)\in N}\left[\kappa_{z,i}\right]. \tag{3}$$

We have the following lemma:

▶ **Lemma 4.7.** *For every $(z,i)\in N$, $\kappa_{z,i}\geq(1-w)\log_2(\frac{1}{3/n+q\epsilon})-O(1)$.*

Proof of Lemma 4.7. Fix $(z,i)\in N$. Let $T^{(=1)}\subseteq T$ be the set of all transcripts that declare the input pair to be at distance 1. Using the extended Gibbs' inequality (Fact 3.6),

$$\kappa_{z,i}=\sum_{t\in T^{(=1)}}\Pr[\Pi(z\oplus e_i,z)=t]\log\frac{\Pr[\Pi(z\oplus e_i,z)=t]}{\mathbb{E}_{j,\ell\in[n]}\Pr[\Pi(z\oplus e_i\oplus e_j,z\oplus e_\ell)=t]}-\ln 2.$$

Using the log-sum inequality, Definition 4.5 and the fact that $x\log_2(x)\geq -0.6$ for all $x\in[0,1]$, we have that

$$\kappa_{z,i}\geq(1-w)\log_2(\frac{1-w}{3/n+q\epsilon})-\ln 2=(1-w)\log_2(\frac{1}{3/n+q\epsilon})-O(1).$$ ◀

Using Lemma 4.7 and Equation (3), we get

$$I((X,Y);\Pi(X,Y)|Z,D,B=0) \geq \frac{|N|}{n2^n}\frac{(1-w)}{2}\log_2(\frac{1}{3/n+q\epsilon}) - O(1)$$
$$\geq (1-\frac{4}{q}-\frac{4(\delta+\epsilon)}{w})\frac{(1-w)}{2}\log_2(\frac{1}{3/n+q\epsilon}) - O(1),$$

where the last inequality follows from Lemma 4.6. The second part of Lemma 4.3 follows from that the fact that

$$I((X,Y);\Pi(X,Y)|Z,D,B) = \frac{1}{2}\Big(I((X,Y);\Pi(X,Y)|Z,D,B=0) + I((X,Y);\Pi(X,Y)|Z,D,B=1)\Big).$$ ◀

5 Upper Bounds on the Complexity of Hamming Distance

5.1 Information Complexity Upper Bound

Algorithm 1 Protocol for $\text{HAM}_{n,d}$

Input. Alice is given $x \in \{0,1\}^n$ and Bob is given $y \in \{0,1\}^n$.
Parameters. $\epsilon \geq 0$, shared random string r.
Output. $\text{HAM}_{n,d}(x,y)$.

1: Alice and Bob use r to define a random k-partition P of $[n]$.
2: Alice sets $a \leftarrow h_P(x)$.
3: Bob sets $b \leftarrow h_P(y)$.
4: Alice and Bob initialize $c = 0$.
5: **for** $i = 1, \ldots, n$ **do**
6: Alice and Bob exchange a_i and b_i.
7: If $a_i \neq b_i$, they both update $c \leftarrow c + 1$.
8: If $c > d$, **return** 0.
9: **end for**
10: **return** 1.

In this section, we describe and analyze the protocol that establishes the upper bound on the information complexity of $\text{HAM}_{n,d}$ stated in Proposition 1.1. The protocol is described in Protocol 1. The analysis of the protocol relies on some basic inequalities that follow from a simple balls-and-bins lemma.

▶ **Definition 5.1** (Dot product). The *dot product* between vectors in $\{0,1\}^n$ is defined by setting $x \cdot y = \sum_{i=1}^n x_i y_i \pmod 2$.

▶ **Definition 5.2** (Random partition). For any $k < n$, a *random k-partition* P of $[n]$ is obtained by defining k sets $S_1, \ldots, S_k$ and putting each element $i \in [n]$ in one of those sets independently and uniformly at random. For $k \geq n$, we simply define P to be the complete partition $\{1\}, \ldots, \{n\}$ of $[n]$. We associate the partition P with a family of k elements $\alpha_1, \ldots, \alpha_k$ in $\{0,1\}^n$ by setting the ith coordinate of α_j to 1 iff $i \in S_j$.

▶ **Definition 5.3** (Hashing operator). For any $k \leq n$, the *k-hashing operator* $h_P : \{0,1\}^n \to \{0,1\}^k$ corresponding to the partition $P = (\alpha_1, \ldots, \alpha_k)$ of $[n]$ is the map defined by $h_P : x \mapsto (x \cdot \alpha_1, \ldots, x \cdot \alpha_k)$.

▶ **Lemma 5.4.** *Fix $d \geq 1$. If we throw at least $d+1$ balls into $(d+2)^2/\delta$ buckets independently and uniformly at random, then the probability that at most d buckets contain an odd number of balls is bounded above by δ.*

Proof. Toss the balls one at a time until the number r of remaining balls and the number t of buckets that contain an odd number of balls satisfy $r + t \leq d + 2$. If we toss all the balls without this condition being satisfied, then in the end we have more than $d + 2 > d + 1$ buckets with an odd number of balls and the lemma holds. Otherwise, fix r, t be the values when the condition $r + t \leq d + 2$ is first satisfied. Since r decreases by 1 everytime we toss a ball and t can only go up or down by 1 for each ball tossed, and since originally $r \geq d + 1$, we have $d + 1 \leq r + t \leq d + 2$. This implies that $r \leq d + 2$, that $t \leq d + 2$ and that if each of the r remaining balls land in one of the $(d+2)^2/\delta - t$ buckets that currently contain an even number of balls, the conclusion of the lemma hold. The probability that this event does not hold is at most

$$\begin{aligned} \frac{t}{(d+2)^2/\delta} + \frac{t+1}{(d+2)^2/\delta} + \cdots + \frac{t+r-1}{(d+2)^2/\delta} &\leq \frac{rt + r(r-1)/2}{(d+2)^2/\delta} \\ &\leq \delta \frac{(\frac{d+2}{2})^2 + (d+2)(d+1)/2}{(d+2)^2} \leq \delta\,. \end{aligned}$$ ◀

▶ **Corollary 5.5.** *For every $x, y \in \{0,1\}^n$, the hashes $a = h_P(x)$ and $b = h_P(y)$ corresponding to a random $((d+2)^2/\epsilon)$-partition P of $[n]$ satisfy $\textsc{Ham}_{n,d}(a,b) = \textsc{Ham}_{n,d}(x,y)$ with probability at least $1 - \epsilon$.*

Proof. Let $S \subseteq [n]$ denote the set of coordinates $i \in [n]$ on which $x_i \neq y_i$. The number of coordinates $j \in [(d+2)^2/\epsilon]$ on which $a_j \neq b_j$ corresponds to the number of parts of the random partition P that receive an odd number of coordinates from S. This number corresponds to the number of buckets that receive an odd number of balls when $|S|$ balls are thrown uniformly and independently at random. When $|S| \leq d$, at most d buckets can contain a ball (and thus an odd number of balls) and so the corollary always holds. When $|S| \geq d + 1$, then by Lemma 5.4, the number of parts with an odd number of is also at least $d + 1$ except with probability at most ϵ. ◀

We are now ready to complete the proof of Proposition 1.1.

Proof of Proposition 1.1. Let us first examine the correctness of the protocol. When $\epsilon < n/(d+2)^2$, the protocol never errs since the players output 1 only when they verify (deterministically) that their strings have Hamming distance at most d. When $\epsilon \geq n/(d+2)^2$, the protocol is always correct when $\textsc{Ham}_{(d+2)^2/\epsilon,d}(a,b) = \textsc{Ham}_{n,d}(x,y)$. This identity always holds when the Hamming distance of x and y is at most d. And when the Hamming distance of x and y is greater than d, the identity is satisfied with probability at least $1 - \epsilon$ by Corollary 5.4.

Let us now analyze the information cost of the protocol. Write $m = \min\{n, (d+2)^2/\epsilon\}$ to denote the length of the vectors a and b. Let $\Pi(x,y)$ denote the transcript of the protocol on inputs x, y. Let μ be any distribution on $\{0,1\}^n \times \{0,1\}^n$. Let (X,Y) be drawn from μ and define $A = h_P(X)$, $B = h_P(Y)$. By the data processing inequality, since $I(\Pi(X,Y); X \mid Y, A) = 0$, the mutual information of $\Pi(X,Y)$ and X given Y satisfies

$$I(\Pi(X,Y); X \mid Y) \leq I(\Pi(X,Y); A \mid Y) = I(\Pi(A,B); A \mid B).$$

Furthermore, with $d \log m$ bits we can identify the first d coordinates $i \in [m]$ for which $a_i \neq b_i$ and thereby completely determine $\Pi(A, B)$. So by Fact 2.1,

$$H(\Pi(X, Y) \mid Y) \leq d \log m.$$

The same argument also yields $I(\Pi(X, Y); Y \mid X) \leq d \log m$, showing that the information cost of the protocol is at most $2d \log m$. ◀

5.2 Communication Complexity

Huang et al. [13], building on previous results by Yao [26] and by Gavinsky et al. [11], showed that the randomized communication complexity of $\textsc{Ham}_{n,d}$ in the simultaneous message passing (SMP) model is bounded above by $R^{\|,\text{pub}}_{1/3}(\textsc{Ham}_{n,d}) = O(d \log d)$. We simplify their protocol and refine this analysis to give a general upper bound on the communication complexity for arbitrary values of ϵ.

▶ **Theorem 5.6.** *Fix $\epsilon > 0$. The randomized communication complexity of $\textsc{Ham}_{n,d}$ in the simultaneous message passing model is bounded above by*

$$R^{\|,\text{pub}}_{\epsilon}(\textsc{Ham}_{n,d}) = O(\min\{d \log n + \log 1/\epsilon, d \log d/\epsilon).$$

The proof of the theorem uses the following results.

▶ **Lemma 5.7.** $R^{\|,\text{pub}}_{\epsilon}(\textsc{Ham}_{n,d}) = O(d \log n + \log 1/\epsilon)$.

Proof. Alice and Bob can generate $q = \log \binom{n}{\leq d} + \log \frac{1}{\epsilon}$ random vectors $r_1, \ldots, r_q \in \{0,1\}^n$ and send the dot products $x \cdot r_1, \ldots, x \cdot r_q$ and $y \cdot r_1, \ldots, y \cdot r_q$ to the verifier, respectively. The verifier then returns 1 iff there is a vector $z \in \{0,1\}^n$ of Hamming weight at most d such that $x \cdot r_j = y \cdot r_j \oplus z \cdot r_j$ for every $j \in [q]$. When $\textsc{Ham}_(x, y) \leq d$, the verifier always returns 1 since in this case $x \cdot r_j = (y \oplus z) \cdot r_j = y \cdot r_j \oplus z \cdot r_j$ for some vector z of Hamming weight at most d. And for any $z \in \{0,1\}^n$, when $x \neq y \oplus z$, the probability that the identity $x \cdot r_j = y \cdot r_j \oplus z \cdot r_j$ holds for every $j \in [q]$ is 2^{-q}. So, by the union bound, the overall probability that the verifier erroneously outputs 1 is at most $\binom{n}{\leq d} 2^{-q} = \epsilon$. ◀

▶ **Lemma 5.8.** $R^{\|,\text{pub}}_{\epsilon}(\textsc{Ham}_{n,d}) \leq R^{\|,\text{pub}}_{\epsilon/2}(\textsc{Ham}_{(d+2)^2/\epsilon,d})$.

Proof. Consider the protocol where Alice and Bob use the shared random string to generate a $(d+2)^2/\epsilon$-hash of their inputs x, y and then apply the protocol for $\textsc{Ham}_{(d+2)^2/\epsilon,d}$ with error $\epsilon/2$. By Corollary 5.5, the probability that the hashed inputs a, b do not satisfy $\textsc{Ham}_{n,d}(a, b) = \textsc{Ham}_{n,d}(x, y)$ is at most $\frac{\epsilon}{2}$. The lemma follows from the union bound. ◀

We can now complete the proof of the theorem.

Proof of Theorem 5.6. When $\epsilon \leq d/n$, Alice and Bob simply run the protocol from the proof of Lemma 5.7. When $\epsilon > d/n$, Alice and Bob combine the protocol from the proof of Lemma 5.8 with the protocol from Lemma 5.7 (with the parameter n set to $(d+2)^2/\epsilon$). ◀

6 Applications and Extensions

6.1 Property Testing Lower Bounds

A Boolean property P is a subset of the set of functions mapping $\{0,1\}^n$ to $\{0,1\}$. A function f has property P if $f \in P$. Conversely, we say that the function f is ϵ-*far* from P if

$|\{x \in \{0,1\}^n : f(x) \neq g(x)\}| \geq \epsilon 2^n$ for every $g \in P$. A (q, ϵ, δ)-*tester* for P is a randomized algorithm A that, given oracle access to some function $f : \{0,1\}^n \to \{0,1\}$, queries the value of f on at most q elements from $\{0,1\}^n$ and satisfies two conditions:

1. When f has property P, A accepts f with probability at least $1 - \delta$.
2. When f is ϵ-far from P, A rejects f with probability at least $1 - \delta$.

The query complexity of the property P for given ϵ and δ parameters is the minimum value of q for which there is a (q, ϵ, δ)-tester for P. We denote this query complexity by $Q_{\epsilon,\delta}(P)$.

The two properties we consider in this section are k-linearity and k-juntas. The function f is *k-linear* iff it is of the form $f : x \mapsto \sum_{i \in S} x_i \pmod 2$ for some set $S \subseteq [n]$ of size $|S| = k$. (The k-linear functions are also known as *k-parity* functions.) The function f is a *k-junta* if there is a set $J = \{j_1, \ldots, j_k\} \subseteq [n]$ of coordinates such that the value of $f(x)$ is determined by the values of $x_{j_1}, \ldots, x_{j_k}$ for every $x \in \{0,1\}^n$.

The upper bound in Corollary 1.9 is from [3]. The proof is obtained via a simple reduction from the Hamming distance function, following the method introduced in [4].

▶ **Corollary 6.1** (Unconditional lower bound of Corollary 1.9). *Fix $0 < \delta < \frac{1}{3}$, $0 < \epsilon \leq \frac{1}{2}$, and $k \leq n/\log \frac{1}{\delta}$. Then $Q_{\epsilon,\delta}(k\text{-Linearity}) = \Omega(k \log \frac{1}{\delta})$ and $Q_{\epsilon,\delta}(k\text{-Juntas}) = \Omega(k \log \frac{1}{\delta})$.*

Proof. Consider the following protocol for the $\textsc{Ham}_{n,k}$ function. Alice takes her input $x \in \{0,1\}^n$ and builds the function $\chi_A : \{0,1\}^n \to \{0,1\}$ defined by $\chi_A : z \mapsto \sum_{i=1}^n x_i z_i \pmod 2$. Similarly, Bob builds the function χ_B from his input y by setting $\chi_B : z \mapsto \sum_{i=1}^n y_i z_i \pmod 2$. Notice that the bitwise XOR of the functions χ_A and χ_B satisfies

$$\chi_A \oplus \chi_B : z \mapsto \sum_{i=1}^{n} (x_i + y_i) z_i \pmod 2 = \sum_{i \in [n] : x_i \neq y_i} z_i \pmod 2.$$

The function $\psi := \chi_A \oplus \chi_B$ is ℓ-linear, where ℓ is the Hamming distance of x and y. When $\ell \leq k$, the function ψ is a k-junta; when $\ell > k$, then ψ is $\frac{1}{2}$-far from all k-juntas. Let Alice and Bob simulate a q-query tester for k-juntas on ψ by exchanging the values of $\chi_A(z)$ and $\chi_B(z)$ for every query z of the tester. If this tester succeeds with probability $1 - \delta$, the resulting protocol is a δ-error protocol for $\textsc{Ham}_{n,k}$ with communication cost at most $2q$. Therefore, by Theorem 1.2, $Q_{\epsilon,\delta}(k\text{-Juntas}) \geq R^{\text{pub}}_{\delta}(\textsc{Ham}_{n,k}) \geq \Omega(k \log \frac{1}{\delta})$.

The lower bound for $Q_{\epsilon,\delta}(k\text{-Linearity})$ is essentially the same except that we use the extra fact that the bound in Theorem 1.2 also holds even when we have the additional promise that the Hamming distance between x and y is either exactly d or greater than d. ◀

The proof of the conditional lower bounds of Corollary 1.9 is identical except that we appeal to the bound in Theorem 1.6 instead of the one in Theorem 1.2 in the conclusion of the proof.

6.2 Parity Decision Tree Complexity Lower Bounds

The proof of Corollary 1.10 is similar to the one in the last section. The details follow.

Proof of Corollary 1.10. Consider the following protocol for the $\textsc{Ham}_{n,d}$ function. Let $z = x \oplus y \in \{0,1\}^n$ denote the bitwise XOR of Alice's input x and Bob's input y. The Hamming weight of z is exactly the Hamming distance between x and y. Recall that a randomized parity decision tree of depth d is a distribution over deterministic parity decision trees that each have depth at most d. Alice and Bob can use their shared randomness to draw a tree T from this distribution. Since for every $S \subseteq [n]$, the parity of z on S, denoted z_S, satisfies $z_S = x_S \oplus y_S$, Alice and Bob can determine the path of z through T by exchanging

the parities x_S and y_S for each query of the parity of z on the subset $S \subseteq [n]$ of coordinates. So they can determine the value of $\textsc{Ham}_{n,d}$ with error at most ϵ using $2R_\epsilon^{\oplus}(\textsc{Weight}_{n,d})$ bits of communication. The bounds in Corollary 1.10 follow directly from Theorems 1.2 and 1.6. ◀

6.3 Symmetric XOR Functions

The key to the proof of Theorem 1.11 is the observation that the proof of Theorem 1.2 proves an even stronger statement: it shows that the same information complexity bound also holds for the $\textsc{Ham}_{n,d\text{vs.}d+2}$ promise version of the $\textsc{Ham}_{n,d}$ function.

▶ **Theorem 6.2** (Strengthening of Theorem 1.2). *For every $1 \leq d < n-1$ and every $0 \leq \epsilon < 1/2$,*

$$\mathrm{IC}(\textsc{Ham}_{n,d\text{vs.}d+2}, \epsilon) = \Omega(\min\{\log \tbinom{n}{d}, d \log(1/\epsilon)\}).$$

Proof. The proof is identical to that of Theorem 1.2. The only additional observation that we need to make is that in our argument, our choice of μ^k ensures that we only ever examine the behavior of the protocol on inputs of the $\mathrm{AND}_d \circ \textsc{Ham}_{n,1\text{vs.}3}$ function in which at most 1 of the d inputs to the $\textsc{Ham}_{n,1\text{vs.}3}$ function have Hamming weight 3. ◀

The proof of Theorem 1.11 follows immediately from Theorem 6.2.

Proof of Theorem 1.11. Consider any ϵ-error protocol P for the symmetric XOR function f. Let $d = \Gamma_{+2}(f)$. Then since $f(d) \neq f(d+2)$, P must distinguish between the cases where Alice and Bob's inputs have Hamming distance d from those where their inputs have Hamming distance $d+2$. Thus, the protocol P (or the protocol P' obtained by flipping the outputs of P) is an ϵ-error protocol for $\textsc{Ham}_{n,d\text{vs.}d+2}$ and so it must have information cost at least $\mathrm{IC}(\textsc{Ham}_{n,d\text{vs.}d+2}, \epsilon)$ and the bound follows from Theorem 6.2. ◀

7 Direct-sum Theorems for Hamming Distance

It was shown in [10] that, when the error rate is viewed as a parameter, the equality function violates the direct-sum theorem for randomized communication complexity in the following sense:

▶ **Definition 7.1.** We say that a function $f : \{0,1\}^m \times \{0,1\}^m \to \{0,1\}$ violates the direct-sum theorem for randomized communication complexity if

$$R_\epsilon^k(f^k) = o(kR_\epsilon(f))$$

where $R_\epsilon^k(f^k)$ denotes the randomized communication complexity of computing f such that on each tuple of k input pairs, the error probability on each input pair is at most ϵ.

Braverman [5] showed that his constant upper bound on the information complexity of EQ (which holds for any error rate $\epsilon \geq 0$) implies a different proof of the fact that EQ violates the direct-sum theorem for randomized communication complexity when $\epsilon = o(1)$ is viewed as a parameter. We next observe that our tight characterization of the information complexity of HD_1^m given in Proposition 1.1 and Theorem 1.2 implies that HD_1^m satisfies the direct-sum theorem for randomized communication complexity whenever $m = \Omega(\mathrm{poly}(1/\epsilon))$ and violates it otherwise (i.e., when $\log m = o(\log(1/\epsilon))$). This can be seen as a further indication of the qualitative difference between the information complexity of EQ and that of HD_1^m in the small error regime.

▶ **Proposition 7.2.** *HD_1^m satisfies the direct-sum theorem for randomized communication complexity whenever $m = \Omega(\text{poly}(1/\epsilon))$ and violates it otherwise (i.e., when $\log m = o(\log(1/\epsilon))$).*

Proof. We first recall the following theorem of Braverman [5]:

▶ **Theorem 7.3** ([5])**.** *For any function f and any error rate $\epsilon > 0$, $IC(f, \epsilon) = \lim_{k\to\infty} \frac{R_\epsilon^k(f^k)}{k}$.*

Applying Theorem 7.3 with $f = HD_1^m$, we get that that $R_\epsilon^k((HD_1^m)^k) = \Theta(kIC(HD_1^m, \epsilon))$. Proposition 1.1 and Theorem 1.2, we have that $IC(HD_1^m, \epsilon) = \Theta(\min(\log m, \log(1/\epsilon)))$. Hence, we get that

$$R_\epsilon^k((HD_1^m)^k) = \Theta(k \min(\log m, \log(1/\epsilon)))$$

On the other hand, we have that $R_\epsilon(HD_1^m) = \Omega(\log(1/\epsilon))$ [5]. So we conclude that

$$R_\epsilon^k((HD_1^m)^k) = \Theta(kR_\epsilon(HD_1^m))$$

whenever $m = \Omega(\text{poly}(1/\epsilon))$ and

$$R_\epsilon^k((HD_1^m)^k) = o(kR_\epsilon(HD_1^m))$$

whenever $\log m = o(\log(1/\epsilon))$. ◀

8 Low Information Protocols for $AND_k \circ \text{Ham}_{n/k,1}$ and $OR_k \circ \text{Ham}_{n/d,1}$

In this section, we give protocols for $AND_k \circ \text{HAM}_{n/k,1}$ and $OR_k \circ \text{HAM}_{n/k,1}$ with $O(k)$ information cost. For $AND_k \circ \text{HAM}_{n/k,1}$, the following theorem implies a protocol with $O(k)$ information cost for any constant error parameter $\epsilon > 0$.

▶ **Theorem 8.1.** *For any error parameter $\epsilon > 0$,*

$$\text{IC}(AND_k \circ \text{HAM}_{n/k,1}, \epsilon) = O(k \min(\log(n/k), \log(1/\epsilon))).$$

Proof. The description of the protocol is given below.

Algorithm 2 Protocol for $AND_k \circ \text{HAM}_{n/k,1}$

Input. Alice is given $x \in \{0,1\}^n$ and Bob is given $y \in \{0,1\}^n$

Output. $AND_k \circ \text{HAM}_{n/k,1}(x, y)$

1: Run in parallel k copies of Algorithm 1 for $\text{HAM}_{n/k,1}$ with error parameter ϵ on $(x^{(1)}, y^{(1)}), \ldots, (x^{(k)}, y^{(k)})$.
2: Declare $AND_k \circ \text{HAM}_{n/k,1}(x, y)$ to be 1 if and only if all the $(x^{(i)}, y^{(i)})$'s were declared to be at distance 1.

If $AND_k \circ \text{HAM}_{n/k,1}(x, y) = 1$, then all the $(x^{(i)}, y^{(i)})$'s are at distance 1. Since Algorithm 1 for $\text{HAM}_{n/k,1}$ always outputs the correct answer on distance-1 input pairs, each $(x^{(i)}, y^{(i)})$ will be declared to be at distance 1 and hence the above protocol will output the correct answer for $AND_k \circ \text{HAM}_{n/k,1}(x, y)$ (namely, 1) with probability 1. If $AND_k \circ \text{HAM}_{n/k,1}(x, y) = 0$, then there exists an $(x^{(i)}, y^{(i)})$ that is at distance 3. Then, the copy of Algorithm 1 for $\text{HAM}_{n/k,1}$

[5] This follows from the fact that $R_\epsilon(EQ) = \Omega(\log(1/\epsilon))$ and by padding.

running on $(x^{(i)}, y^{(i)})$ will declare this pair to be at distance 3 with probability at least $1-\epsilon$. Thus, the above protocol will output the correct answer for $AND_k \circ \text{HAM}_{n/k,1}(x,y)$ (namely, 0) with probability at least $1-\epsilon$. Fix a distribution μ on the input pair (X,Y) with support $\{0,1\}^{2n}$ and let $\mu^{(i)}$ denote the marginal of μ over $(X^{(i)}, Y^{(i)})$ for every $i \in [k]$. Denoting by Π the transcript of the above protocol, its information cost $\text{IC}_\mu(\Pi) := I_\mu(\Pi; X|Y) + I_\mu(\Pi; Y|X)$ is upper-bounded by the following lemma:

▶ **Lemma 8.2.** $\text{IC}_\mu(\Pi) = O(k \min(\log(n/k), \log(1/\epsilon)))$.

Proof. Denote by $\Pi^{(1)}, \ldots, \Pi^{(k)}$ the transcripts corresponding to the k parallel runs of Algorithm 1 for $\text{HAM}_{n/k,1}$ on the input pairs $(x^{(1)}, y^{(1)}), \ldots, (x^{(k)}, y^{(k)})$ respectively. Since $\Pi^{(1)}, \ldots, \Pi^{(k)}$ completely determine Π, we have that

$$\text{IC}_\mu(\Pi) = I_\mu(\Pi^{(1)}, \ldots, \Pi^{(k)}; X|Y) + I_\mu(\Pi^{(1)}, \ldots, \Pi^{(k)}; Y|X).$$

Since each of the protocols $\Pi^{(1)}, \ldots, \Pi^{(k)}$ - as well as Π - is completely symmetric with respect to Alice and Bob, it is enough to show that $I_\mu(\Pi^{(1)}, \ldots, \Pi^{(k)}; X|Y) = O(k \min(\log(n/k), \log(1/\epsilon)))$. By the chain rule for mutual information, we have that:

$$\begin{aligned}
I_\mu(\Pi^{(1)}, \ldots, \Pi^{(k)}; X|Y) &= \sum_{i=1}^{k} I_\mu(\Pi^{(i)}; X|Y, \Pi^{(<i)}) \\
&= \sum_{i=1}^{k} \sum_{j=1}^{k} I_\mu(\Pi^{(i)}; X^{(j)}|Y, \Pi^{(<i)}, X^{(<j)}) \\
&\overset{(a)}{=} \sum_{i=1}^{k} I_\mu(\Pi^{(i)}; X^{(i)}|Y, \Pi^{(<i)}, X^{(<i)}) \\
&\overset{(b)}{=} \sum_{i=1}^{k} I_{\mu^{(i)}}(\Pi^{(i)}; X^{(i)}|Y^{(i)}) \\
&\overset{(c)}{=} \sum_{i=1}^{k} O(\min(\log(n/k), \log(1/\epsilon))) \\
&= O(k \min(\log(n/k), \log(1/\epsilon)))
\end{aligned}$$

where (a) follows from $\Pi^{(i)}$ and $X^{(j)}$ being conditionally independent given $Y, \Pi^{(<i)}, X^{(<j)}$ for any $i \neq j \in [k]$, (b) follows from $(\Pi^{(i)}, X^{(i)})$ being conditionally independent of $Y^{(\neq i)}, \Pi^{(<i)}, X^{(<i)}$ given $Y^{(i)}$ and (c) follows from Proposition 1.1. ◀

The previous lemma implies that for constant ϵ, the information cost of protocol Π is $O(k)$. The following lemma notes that, in this case, even the communication complexity is $O(k)$:

▶ **Lemma 8.3.** *For constant ϵ, the communication complexity of Algorithm 2 is $O(k)$.*

Proof. Note that for constant ϵ, Theorem 5.6 implies that each run of Algorithm 1 has communication cost $O(1)$. Since Algorithm 2 performs k such calls to Algorithm 1, the communication cost of Algorithm 2 is hence $O(k)$. ◀

◀

▶ **Theorem 8.4.** *For every constant $\nu \in (0,1)$, $CC(OR_k \circ \text{HAM}_{n/k,1}, 1/k^\nu) = O(k)$.*

Algorithm 3 Algorithm for $OR_k \circ \text{HAM}_{n/k,1}$

Input. Alice is given $x \in \{0,1\}^n$ and Bob is given $y \in \{0,1\}^n$
Output. $OR_k \circ \text{HAM}_{n/k,1}(x,y)$

1: Let $c := \nu + 1$, $\eta := 1/4$, $t := c \log_2 k$, and $h := t/2$.
2: Mark all k input pairs $(x^{(1)}, y^{(1)}), \ldots, (x^{(k)}, y^{(k)})$ as distance-1 pairs.
3: Initialize the number u of inputs pairs that are marked to be at distance 1: $u = k$.
4: **for** $i = 1 : t$ **do**
5: Run in parallel u copies of Protocol 1 for $\text{HAM}_{n/k,1}$ with error parameter $\epsilon' = 1/2$ on each of the input pairs $(x^{(i)}, y^{(i)})$ that are still marked as distance-1 pairs.
6: If an input pair is declared to be at distance 3, mark it as a distance-3 pair.
7: If $i \leq h$ and the number u of input pairs that are still marked as distance-1 pairs is larger than $(1+\eta)k/2^i$, halt and declare $OR_k \circ \text{HAM}_{n/k,1}(x,y)$ to be 1.
8: **end for**
9: Declare $OR_k \circ \text{HAM}_{n/k,1}(x,y)$ to be 0 if and only if all the $(x^{(i)}, y^{(i)})$'s are marked as distance-3 pairs.

Proof. The description of the protocol is given in Algorithm 3.

If $OR_k \circ \text{HAM}_{n/k,1}(x,y) = 1$, then there is an input pair $(x^{(i)}, y^{(i)})$ (for some $i \in [k]$) that is at distance 1. Since Protocol 1 for $\text{HAM}_{n/k,1}$ always outputs the correct answer on distance-1 input pairs, $(x^{(i)}, y^{(i)})$ will be declared to be at distance 1 in each iteration and hence the above protocol will output the correct answer for $OR_k \circ \text{HAM}_{n/k,1}(x,y)$ (namely, 1) with probability 1. If $OR_k \circ \text{HAM}_{n/k,1}(x,y) = 0$, then the protocol outputs the correct answer with probablity at least $1 - \epsilon$ as shown by the following lemma:

▶ **Lemma 8.5.** *If $OR_k \circ \text{HAM}_{n/k,1}(x,y) = 0$, then the probability that the protocol outputs a wrong answer is at most $1/k^{c-1} + ke^{-\frac{\eta^2 k^{1-c/2}}{3}}$.*

Proof. If $OR_k \circ \text{HAM}_{n/k,1}(x,y) = 0$, all the $(x^{(i)}, y^{(i)})$'s are at distance 3. Conditioned on the fact that the protocol didn't halt and output 1 during the for loop, the probability that the protocol outputs an incorrect answer is, by the union bound, at most $k \times 1/2^t = 1/k^{c-1}$. To complete the proof, we now upper bound the probability that the protocol halts and outputs 1 during the for loop. Note that the expected number of input pairs that are marked as distance-1 pairs after the i-th iteration is $k/2^i$. By the Chernoff bound, the probability that after the i-th iteration, the number of distance-1 marked pairs is larger than $(1+\eta)k/2^i$ is at most

$$e^{-\eta^2 k/(3\times 2^i)} \leq e^{-\eta^2 k/(3\times 2^h)} = e^{-\frac{\eta^2 k^{1-c/2}}{3}}.$$

By the union bound, the probability that the algorithm halts and outputs 0 during the for loop is at most $ke^{-\frac{\eta^2 k^{1-c/2}}{3}}$. By another union bound, the probability that the protocol outputs an incorrect answer is at most $1/k^{c-1} + ke^{-\frac{\eta^2 k^{1-c/2}}{3}}$. ◀

▶ **Lemma 8.6.** *For any constant $c \in (1,2)$, the communication complexity of the above protocol is $O(1)$.*

Proof. Consider the execution of Protocol 3. For every $i \in [h]$, the number of calls to Protocol 1 is at most $k(1+\eta)/2^{i-1}$. For every $i \in \{h+1, \ldots, k\}$, the number of calls of

Protocol 1 is at most $k(1+\eta)/2^h$. Hence, the total number of calls to Protocol 1 is at most:

$$\sum_{i=1}^{h} \frac{k(1+\eta)}{2^{i-1}} + \frac{hk(1+\eta)}{2^h} \leq 2k(1+\eta) + \frac{ck(1+\eta)\log_2 k}{2^{\frac{ck\log_2 k}{2}+1}} = 2k(1+\eta) + \frac{c(1+\eta)}{2} k^{1-c/2}\log_2 k = \Theta(k)$$

where the last equality uses the fact that $c \in (1,2)$ is a constant. By Theorem 5.6, the communication cost of any run of Protocol 1 with noise rate $\epsilon' = 1/2$ is $O(1)$. Hence, the communication cost of Protocol 3 is $O(1)$. ◀

Using Lemma 8.5 (and the paragraph preceding it), Lemma 8.6 and the fact that $\nu = c-1$ is a constant in $(0,1)$, the statement of Theorem 8.4 now follows. ◀

Acknowledgments. The authors would like to thank Madhu Sudan for very helpful discussions. They also wish to thank the anonymous referees for much valuable feedback.
Eric Blais is supported by a Simons Postdoctoral Fellowship.

References

1 Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. In *Proc. 43rd Annual IEEE Symposium on Foundations of Computer Science*, pages 209–218, 2002.

2 Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In *STOC*, pages 67–76, 2010.

3 Eric Blais. Testing juntas nearly optimally. In *Proceedings of the 41st annual ACM symposium on Theory of computing*, pages 151–158. ACM, 2009.

4 Eric Blais, Joshua Brody, and Kevin Matulef. Property testing lower bounds via communication complexity. *Computational Complexity*, 21(2):311–358, 2012.

5 Mark Braverman. Interactive information complexity. In *Proc. 44th Annual ACM Symposium on the Theory of Computing*, 2012.

6 Mark Braverman and Anup Rao. Information equals amortized communication. In *FOCS*, pages 748–757, 2011.

7 Amit Chakrabarti and Oded Regev. An optimal lower bound on the communication complexity of Gap-Hamming-Distance. *SIAM Journal on Computing*, 41(5):1299–1317, 2012.

8 Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on*, pages 270–278. IEEE, 2001.

9 Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

10 Tomas Feder, Eyal Kushilevitz, Moni Naor, and Noam Nisan. Amortized communication complexity. *SIAM Journal on Computing*, 24(4):736–750, 1995.

11 Dmitry Gavinsky, Julia Kempe, and Ronald de Wolf. Quantum communication cannot simulate a public coin. *arXiv preprint quant-ph/0411051*, 2004.

12 Johan Håstad and Avi Wigderson. The randomized communication complexity of set disjointness. *Theory of Computing*, 3(1):211–219, 2007.

13 Wei Huang, Yaoyun Shi, Shengyu Zhang, and Yufan Zhu. The communication complexity of the hamming distance problem. *Inform. Process. Lett.*, 99:149–153, 2006.

14 Bala Kalyanasundaram and Georg Schnitger. The probabilistic communication complexity of set intersection. *SIAM J. Disc. Math.*, 5(4):547–557, 1992.

15 Iordanis Kerenidis, Sophie Laplante, Virginie Lerays, Jérémie Roland, and David Xiao. Lower bounds on information complexity via zero-communication protocols and applications. In *Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium on*, pages 500–509. IEEE, 2012.

16 Marco Molinaro, David P Woodruff, and Grigory Yaroslavtsev. Beating the direct sum theorem in communication complexity with implications for sketching. In *SODA*, pages 1738–1756. SIAM, 2013.

17 Ryan O'Donnell. Hardness amplification within NP. *J. Comput. Syst. Sci.*, 69(1):68–94, 2004.

18 King F Pang and Abbas El Gamal. Communication complexity of computing the hamming distance. *SIAM Journal on Computing*, 15(4):932–947, 1986.

19 Ramamohan Paturi. On the degree of polynomials that approximate symmetric boolean functions (preliminary version). In *STOC*, pages 468–474, 1992.

20 Mert Sağlam and Gábor Tardos. On the communication complexity of sparse set disjointness and exists-equal problems. In *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*, pages 678–687. IEEE, 2013.

21 Alexander A Sherstov. The communication complexity of gap hamming distance. *Theory of Computing*, 8(1):197–208, 2012.

22 Thomas Vidick. A concentration inequality for the overlap of a vector on a large set, with application to the communication complexity of the gap-hamming-distance problem. *Chicago Journal of Theoretical Computer Science*, 1, 2012.

23 Emanuele Viola and Avi Wigderson. Norms, XOR lemmas, and lower bounds for polynomials and protocols. *Theory of Computing*, 4(1):137–168, 2008.

24 David P Woodruff and Qin Zhang. Tight bounds for distributed functional monitoring. In *Proceedings of the 44th symposium on Theory of Computing*, pages 941–960. ACM, 2012.

25 Andrew C. Yao. Some complexity questions related to distributive computing. In *Proc. 11th Annual ACM Symposium on the Theory of Computing*, pages 209–213, 1979.

26 Andrew Chi-Chih Yao. On the power of quantum fingerprinting. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 77–81. ACM, 2003.

A Slight Generalization of the Direct-sum Theorem of [16]

We start by recalling the direct-sum theorem of Molinaro, Woodruff and Yaroslavtsev ([16]), which is stated in terms of the following notion of information complexity:

▶ **Definition 1.1** (MWY notion of information complexity with abortion)**.** Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a function. Then, $\mathrm{IC}_{\mu,\alpha,\delta,\epsilon}(f|\nu)$ is the minimum conditional information cost of a randomized protocol that with probability at least $1-\alpha$ gives a deterministic protocol that computes f with abortion probability at most δ with respect to μ and with conditional error probability given no abortion at most ϵ with respect to μ.

▶ **Theorem 1.2** ([16])**.** *Let $X \in \mathcal{X}$, $Y \in \mathcal{Y}$ and λ be a distribution on (X,Y,D) with marginals μ over (X,Y) and ν over D such that for every value d of D, X and Y are conditionally independent given $D = d$. For any $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, $k \in \mathbb{N}$ and $\delta \leq 1/3$,* $\mathrm{IC}_{\mu^k,\delta}(f^k|\nu^k) = k \cdot \Omega(\mathrm{IC}_{\mu,1/20,1/10,\delta/k}(f|\nu))$

We now give the slight generalization of the MWY theorem that is used to prove Theorem 1.6.

▶ **Theorem 4.2** (Slight generalization of the direct-sum theorem of [16])**.** *Let $X \in \mathcal{X}$, $Y \in \mathcal{Y}$ and λ be a distribution on (X,Y,D) with marginals μ over (X,Y) and ν over D such that*

for every value d of D, X and Y are conditionally independent given $D = d$. For any $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, $k \in \mathbb{N}$ and $\epsilon \leq 1/3$, $\mathrm{IC}_{\mu^k,\epsilon}(f^k|\nu^k) = k \cdot \Omega(\mathrm{IC}_{\mu,O(\epsilon),O(\epsilon/k)}(f|\nu))$.

Proof. For every $i \in [k]$, we denote by W_i the pair (X_i, Y_i) and by $f(W_{<i})$ the tuple $(f(W_1), \ldots, f(W_{i-1}))$.

▶ **Definition 1.3** (Good indices)**.** An index $i \in [k]$ is said to be *good* if

$$\Pr_{\mu,\Pi}[\Pi_i(W) = f(W_i)|\Pi_{<i}(W) = f(W_{<i})] = 1 - O(\epsilon/k)$$

▶ **Lemma 1.4.** *At least half of the indices $i \in [k]$ are good.*

Proof. Follows from averaging and the fact that

$$\Pr_{\mu,\Pi}[\Pi(W) = f(W)] = \prod_{i=1}^{k} \Pr_{\mu,\Pi}[\Pi_i(W) = f(W_i)|\Pi_{<i}(W) = f(W_{<i})] \geq 1 - \epsilon. \quad \blacktriangleleft$$

▶ **Definition 1.5** (Reasonable prefixes)**.** Fix a good index $i \in [k]$. A prefix $w_{<i}$ is said to be *reasonable* if

1. $I(\Pi(W); W|\nu^k, W_{<i} = w_{<i}) = O(I(\Pi(W); W|\nu^k))$
2. $\Pr_{\mu,\Pi}[\Pi_{<i}(W) = f(W_{<i})|W_{<i} = w_{<i}] = 1 - O(\epsilon)$
3. $\Pr_{\mu,\Pi}[\Pi_i(W) = f(W_i)|\Pi_{<i}(W) = f(W_{<i}), W_{<i} = w_{<i}] = 1 - O(\epsilon/k)$

▶ **Lemma 1.6.** *For every good index $i \in [k]$, a random prefix $w_{<i}$ is reasonable with probability at least $1/2$.*

Proof. Follows from 3 applications of Markov's inequality, the union bound and sufficiently large constants in the $O(\cdot)$ notations. ◀

▶ **Definition 1.7** (Acceptable fixings d_{-i})**.** Fix a good index $i \in [k]$ and a reasonable prefix $w_{<i}$. A fixing d_{-i} of D_{-i} is said to be *acceptable* if

1. $I(\Pi(W); W|\nu^k, W_{<i} = w_{<i}, D_{-i} = d_{-i}) = O(I(\Pi(W); W|\nu^k, W_{<i} = w_{<i}))$
2. $\Pr_{\mu,\Pi}[\Pi_{<i}(W) = f(W_{<i})|W_{<i} = w_{<i}, D_{-i} = d_{-i}] = 1 - O(\epsilon)$
3. $\Pr_{\mu,\Pi}[\Pi_i(W) = f(W_i)|\Pi_{<i}(W) = f(W_{<i}), W_{<i} = w_{<i}, D_{-i} = d_{-i}] = 1 - O(\epsilon/k)$

▶ **Lemma 1.8.** *Fix a good index $i \in [k]$ and a reasonable prefix $w_{<i}$. Then, a random fixing d_{-i} of D_{-i} is acceptable with probability at least $1/2$.*

Proof. Follows from 3 applications of Markov's inequality, the union bound and sufficiently large constants in the $O(\cdot)$ notations. ◀

▶ **Lemma 1.9.** *Fix a good index $i \in [k]$, a reasonable prefix $w_{<i}$ and an acceptable fixing d_{-i}. Then, we have that:*

$$I(\Pi(W); W|\nu^k, W_{<i} = w_{<i}, D_{-i} = d_{-i}) \geq \mathrm{IC}_{\mu,O(\epsilon),O(\epsilon/k)}(\mathrm{HAM}_{n,1}|\nu)$$

Proof. The new protocol Π' simulates the old protocol with $W_{<i} = w_{<i}$ and $D_{-i} = d_{-i}$ hardwired and it doesn't use any public randomness beyond that of the old protocol. Hence,

$$I(\Pi(W); W|\nu^k, W_{<i} = w_{<i}, D_{-i} = d_{-i}) \geq I(\Pi'(W_i); W_i|\nu). \quad \blacktriangleleft$$

The lemma now follows from the chain rule for mutual information and Lemmas 1.4, 1.6 and 1.8. ◀