Direct Sum and Partitionability Testing over General Groups

Andrej Bogdanov
Department of Computer Science and Engineering and Institute of Theoretical Computer Science and Communications, Chinese University of Hong Kong, China

Gautam Prakriya
Institute of Theoretical Computer Science and Communications,
Chinese University of Hong Kong, China

Abstract

A function $f(x_1, \ldots, x_n)$ from a product domain $\mathcal{D}_1 \times \cdots \times \mathcal{D}_n$ to an abelian group $\mathcal{G}$ is a *direct sum* if it is of the form $f_1(x_1) + \cdots + f_n(x_n)$. We present a new 4-query direct sum test with optimal (up to constant factors) soundness error. This generalizes a result of Dinur and Golubev (RANDOM 2019) which is tailored to the target group $\mathcal{G} = \mathbb{Z}_2$. As a special case, we obtain an optimal *affinity test* for $\mathcal{G}$-valued functions on domain $\{0,1\}^n$ under product measure. Our analysis relies on the hypercontractivity of the binary erasure channel.

We also study the testability of *function partitionability* over product domains into disjoint components. A $\mathcal{G}$-valued $f(x_1, \ldots, x_n)$ is *k-direct sum partitionable* if it can be written as a sum of functions over k nonempty disjoint sets of inputs. A function $f(x_1, \ldots, x_n)$ with unstructured product range $\mathcal{R}^k$ is *direct product partitionable* if its outputs depend on disjoint sets of inputs.

We show that direct sum partitionability and direct product partitionability are one-sided error testable with $O((n-k)(\log n + 1/\epsilon) + 1/\epsilon)$ adaptive queries and $O((n/\epsilon)\log^2(n/\epsilon))$ nonadaptive queries, respectively. Both bounds are tight up to the logarithmic factors for constant ϵ even with respect to adaptive, two-sided error testers. We also give a non-adaptive one-sided error tester for direct sum partitionability with query complexity $O(kn^2(\log n)^2/\epsilon)$.

2012 ACM Subject Classification Theory of computation → Streaming, sublinear and near linear time algorithms; Theory of computation → Randomness, geometry and discrete structures

Keywords and phrases Direct Sum Test, Function Partitionability, Hypercontractive Inequality, Property Testing

Digital Object Identifier 10.4230/LIPIcs.ICALP.2021.33

Category Track A: Algorithms, Complexity and Games

Related Version *Full Version*: `https://eccc.weizmann.ac.il/report/2020/164/` [5]

Funding *Andrej Bogdanov*: Work funded by Hong Kong RGC GRF grants CUHK14207618 and CUHK14209920.

Acknowledgements We thank Chandra Nair and Yan Nan Wang for valuable discussions on the hypercontractivity of the binary erasure channel and an anonymous reviewer for bringing to our attention the works [2] and [14].

1 Introduction

In their seminal result, Blum, Luby and Rubinfeld [4] gave a four query test to determine whether a function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is affine. We consider a natural generalization of the notion of affinity to functions $f(x_1, \cdots, x_n)$ from $\{0,1\}^n$ to an arbitrary abelian group $\mathcal{G}$: Is f of the form $x_1 \cdot g_1 + \cdots + x_n \cdot g_n + g_0$ for some group elements $g_0, \ldots, g_n \in \mathcal{G}$? The analysis of Blum, Luby and Rubinfeld does not apply unless there is a group homomorphism from the domain to the range.

48th International Colloquium on Automata, Languages, and Programming (ICALP 2021).
Editors: Nikhil Bansal, Emanuela Merelli, and James Worrell; Article No. 33; pp. 33:1–33:19
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

In this work we give an optimal four query affinity test for functions from $\{0,1\}^n$ to an arbitrary abelian group $\mathcal{G}$.

More generally, our test can be used to determine if a function $f(x_1, \dots, x_n)$ from a finite product domain $\mathcal{D}_1 \times \cdots \times \mathcal{D}_n$ to an abelian group $\mathcal{G}$ is a *direct sum*, i.e., whether f is of the form $\sum f_i(x_i)$. This resolves a conjecture of Dinur and Golubev [9].

In contrast to the work of Blum, Luby, and Rubinfeld, which was primarily motivated by applications to probabilistically checkable proofs, direct sum testing over general groups arises in the context of testing *function partionability*: Can a multivariate function be decomposed into independent or loosely related components? Bogdanov and Wang [6] discuss the relevance of this question for real-valued functions to the problem of identifying decompositions of control variables in high-dimensional reinforcement learning. In that setting a direct sum decomposition of the advantage function f describes a system that can be partitioned into independent components, which are lower-dimensional and therefore typically easier to learn. An efficient testing algorithm can be used to probe the existence of such a decomposition before any effort is expended into learning it.

In this work we consider the following two natural partitioning problems for discrete functions over product domains $\mathcal{D}_1 \times \cdots \times \mathcal{D}_n$ (endowed with a product distribution):

- A *direct sum partition* ($\oplus$-partition) of f into k components is a representation of the form $f(x_1, \dots, x_n) = f_1(x_{S_1}) + \cdots + f_k(x_{S_k})$, where $S_1, \dots, S_k$ are disjoint nonempty sets of variables. Here, the range of f is an abelian group $(\mathcal{G}, +)$.
- A *direct product partition* ($\otimes$-partition) of f is a representation of the form $f(x_1, \dots, x_n) = (f_1(x_{S_1}), \dots, f_k(x_{S_k}))$, where $S_1, \dots, S_k$ are disjoint nonempty sets of variables. Here, the range of f is a k-product set $\mathcal{R}^k$.

We are interested in the query complexity of testing partitionability: Given oracle access to f and parameters k, ϵ, how many queries does it take to tell whether f is partitionable or ϵ-far from partitionable?

The related tasks of *direct product testing* and *direct sum testing* ask for the existence of such representations under a known (fixed) partition of inputs. Motivated by applications to probabilistically checkable proofs, Dinur and Steurer [10] and Dickstein and Dinur [8] analyze a 2-query direct product test of essentially optimal soundness.

The query complexity of direct sum testing for $\mathbb{Z}_2$-valued functions, that is of testing whether a function $f\colon \mathcal{D}_1 \times \cdots \times \mathcal{D}_n \to \mathbb{Z}_2$ is of the form $f(x_1, \dots, x_n) = f_1(x_1) + \cdots + f_n(x_n)$, was recently resolved by Dinur and Golubev [9]. They proposed and analysed a 4-query test of optimal (up to constant factors) soundness error. Their tester does not naturally extend to functions valued in arbitrary abelian groups.

Bogdanov and Wang [6] proposed an agnostic learning algorithm for unknown direct sum partitions. As a consequence of their analysis they concluded that $\oplus$-partitionability is testable with $O(kn^3/\epsilon)$ non-adaptive queries. They also showed that $\Omega(n - k + 1)$ queries are necessary for constant ϵ. To the best of our knowledge $\otimes$-partitionability has not been studied before.

Our Results

We analyze a new 4-query direct sum test for functions valued over arbitrary abelian groups. The test is based on the following dual characterization: $f\colon \mathcal{D}_1 \times \cdots \times \mathcal{D}_n \to \mathcal{G}$ is a direct sum $f_1(x_1) + \cdots + f_n(x_n)$ if and only if $D_f(S, \overline{S}; x, y) = 0$ for all pairs of inputs x, y and partitions $(S, \overline{S})$ of $[n]$, where

$$D_f(S, \overline{S}; x, y) = f(x) - f(y_S x) - f(y_{\overline{S}} x) + f(y).$$

Here and in the rest of the manuscript, $y_S x$ is the string in $\mathcal{D}_1 \times \cdots \times \mathcal{D}_n$ that matches y in the S-coordinates and x in the other coordinates. We assume that the domain $\mathcal{D}_1 \times \cdots \times \mathcal{D}_n$ is furnished with a product distribution: For x chosen at random from $\mathcal{D}_1 \times \cdots, \times \mathcal{D}_n$, the coordinates $x_1, \ldots, x_n$ are independent.

The tester accepts if $D_f(S, \overline{S}; x, y) = 0$ for random independent inputs $x, y \in \mathcal{D}_1 \times \cdots \times \mathcal{D}_n$ and a uniformly random partition $(S, \overline{S})$ of $[n]$. Our main result is an optimal (up to constant factor) bound on the soundness error $\rho(f) = \Pr[D_f(S, \overline{S}; x, y) \neq 0]$ of this test in terms of the distance $\delta(f) = \min_g\{\Pr_x[f(x) \neq g(x)] \colon g \text{ is a direct sum}\}$.

▶ **Theorem 1.** *There is an absolute constant $c > 0$ such that for every collection of finite sets $\mathcal{D}_1, \ldots, \mathcal{D}_n$, every abelian group $\mathcal{G}$, and every $f \colon \mathcal{D}_1 \times \cdots \times \mathcal{D}_n \to \mathcal{G}$, $\rho(f) \geq c \cdot \delta(f)$.*

An important special case of the theorem concerns the Boolean domain $\mathcal{D}_1 = \cdots = \mathcal{D}_n = \{0,1\}$ under the uniform distribution (Proposition 5). The class of direct sums from $\{0,1\}^n$ to $\mathcal{G}$ is then precisely the class of affine functions $f(x) = x_1 g_1 + \cdots + x_n g_n + g_0$ for some group elements $g_0, g_1, \ldots, g_n \in \mathcal{G}$.

Using Theorem 1, we obtain the following upper bound on the query complexity of $\oplus$-partitionability.

▶ **Theorem 2.** *Direct sum partitionability over any abelian group is one-sided testable with $O((n-k)(\log n + 1/\epsilon) + 1/\epsilon)$-queries.*

We also prove an upper bound on the query complexity of $\otimes$-partitionability:

▶ **Theorem 3.** *Direct product partitionability is one-sided testable with $O((n/\epsilon)\log^2(n/\epsilon))$ non-adaptive queries.*

The testers in Theorem 2 and 3 are time-efficient.

By the lower bound of Bogdanov and Wang [6], the $\oplus$-partitionability tester is tight up to the $\log n$ factor for constant ϵ. In the special case when $k = n$, direct sum partitionability reduces to direct sum testing and the query complexity is the same as that of Theorem 1.

Our tester for $\oplus$-partitionability is adaptive. We also give a one-sided non-adaptive tester of query complexity $O(kn^2(\log n)^2/\epsilon)$. A non-adaptive lower bound of $\Omega((n-k+1)\,(\log(n-k+1)/\epsilon^c)\log(\log(n-k+1)/\epsilon^c))$ for any $c > 1$ follows from the work of Servedio et al. [13] on junta testing. As in the case of juntas, it follows that adaptivity helps in testing $\oplus$-partitionability for some settings of parameters.

The $\otimes$-partitionability tester is also nearly tight: We show that direct product partitionability requires $\Omega(n)$ queries for $\epsilon = 1/2$ for adaptive testers, and $\Omega(\frac{n}{\epsilon\log(1/\epsilon)})$ queries for non-adaptive testers, for every $k \geq 2$.

The non-adaptive test for $\oplus$-partitionability, and the lower bounds for $\oplus$-partitionability and $\otimes$-partitionability are deferred to the full version of this paper [5].

Ideas and Techniques

Direct sum testing over general groups

The main ingredient of Dinur and Golubev's direct sum tester for $\mathbb{Z}_2$-valued functions is an implicit reduction from general product domains to the Boolean domain $\{0,1\}^n$ under the uniform distribution. We abstract and generalize their reduction. In interest of space we defer the details of this reduction to full version of this paper [5]. To complete their proof, Dinur and Golubev instantiate the reduction with the $\mathbb{Z}_2$-affinity test of Blum, Luby, and Rubinfeld [4].

Table 1 Summary of algorithmic results. All tests have one-sided error.

Property	Our Results	Prior Work
Direct Sum: $f : \mathcal{D}_1 \times \cdots \times \mathcal{D}_n \to \mathcal{G}$ s.t. $f(x) = f(x_1) + \cdots + f(x_n)$	4-query test for arbitrary abelian group $\mathcal{G}$ (Theorem 1)	4-query test for $\mathcal{G} = \mathbb{Z}_2$ [9]
k-$\oplus$-partitionability: $f : \mathcal{D}_1 \times \cdots \times \mathcal{D}_n \to \mathcal{G}$ s.t. $\exists S_1, \ldots, S_k \subseteq [n]$, $f(x) = f(x_{S_1}) + \cdots + f(x_{S_k})$	$O((n-k)(\log n + 1/\epsilon) + 1/\epsilon)$-query adaptive test (Theorem 2) $O(kn^2(\log n)^2/\epsilon)$-query non-adaptive test (See full version [5])	$O(kn^3/\epsilon)$-query non-adaptive test [6]
k-$\otimes$-partitionability: $f : \mathcal{D}_1 \times \cdots \times \mathcal{D}_n \to \mathcal{R}^k$ s.t. $\exists S_1, \ldots, S_k \subseteq [n]$, $f(x) = (f(x_{S_1}), \ldots, f(x_{S_k}))$	$O((n/\epsilon)\log^2(n/\epsilon))$-query non-adaptive test (Theorem 3)	

Our main technical contribution is a tight analysis of the affinity test D_f applied to functions $f\colon \{0,1\}^n \to \mathcal{G}$ valued in an arbitrary abelian group $\mathcal{G}$. To give a sense why the test is sound, let us argue that $\rho(f) = \Omega(\delta(f))$ under the additional assumption that f is close to a direct sum, say if $\delta = \delta(f) \leq 1/27$.

Let B be the set of measure at most $1/27$ on which f differs from its closest direct sum. We claim that conditioned on $x \in B$, the probability that any of the other test queries y, $y_S x$, $y_{\overline{S}} x$ land in B is at most $\delta + 2\delta^{1/3}$. By independence, the probability that $y \in B$ conditioned on $x \in B$ is exactly δ. In contrast, $y_S x$ and $y_{\overline{S}} x$ are not independent of x, but can be sampled by processing x through a binary symmetric channel with crossover probability $1/4$. The bound $\Pr[y_S x \in B | x \in B] \leq \delta^{1/3}$ follows from the small-set expansion of this channel [1], which is equivalent to the hypercontractivity of the corresponding Markov operator [7]. Since the event "$x \in B$ and $y_S x \notin B$ and $y_{\overline{S}} x \notin B$ and $y \notin B$" results in rejection, it follows that $\rho(f) \geq \delta \cdot (1 - \delta - 2\delta^{1/3})$, which is at least $\frac{8}{27}\delta$ by the closeness assumption on f.

For larger values of $\delta(f)$, our proof strategy is to argue that f can be *decoded* to a direct sum function by making at most $O(\rho(f))$ "changes" to the truth-table of f. The decoding algorithm we analyze in Lemma 6 is *iterative plurality* (i.e., iterative maximum likelihood). We show that the function

$$\phi(x) = \text{plurality}_{S,y}\ f(y_S x) + f(y_{\overline{S}} x) - f(y) \tag{1}$$

is, on the one hand, $2\delta(f)$-close to f, and on the other hand, has substantially smaller rejection probability of $\rho(\phi) \leq \rho(f)/2$. By iterating the decoding, i.e. applying the plurality to ϕ again, we arrive at a function that is $4\delta(f)$ close to f and passes the test with probability one, thus must equal a direct product.

This argument is inspired by the linearity test analysis of Blum, Ruby, and Rubinfeld (BLR), who also decode f to a function that is, on the one hand, close to f and, on the other hand, passes their test with probability 1. However, unlike the BLR decoder which yields a linear function after a single round of self-correction, ours inherently requires multiple iterations. For example, if f is a direct sum corrupted on all inputs with relative hamming weight around $1/4$, then $\phi(0)$ is unlikely to be correctly decoded (as $y_S 0$ and $y_{\overline{S}} 0$ will typically be corrupted) and so will typically be inconsistent with a direct sum.

Nevertheless, the high-level structure of our argument closely parallels the BLR analysis. First, in Claim 10 we show that for all but an $o(\rho)$-fraction of inputs x, the plurality in (1) is a strong majority consistent with 99% of the choices of $(S, \overline{S})$ and y. Second, we use the algebraic structure of our test (Claim 9) to show that if $D_\phi(S, \overline{S}; x, y) \neq 0$ then

$D_f(U, V; w, z) \neq 0$ for a substantially larger fraction of query sequences $(w, z_U w, z_{\overline{U}} w, z)$ that can be sampled by applying suitable "noise" to $(S, \overline{S}; x, y)$. If we represent the partition $(S, \overline{S})$ by a binary string $\sigma \in \{0,1\}^n$ (with 1 and 0 indicating memberships in S and $\overline{S}$, respectively), we show that the relevant noise can be modeled by independent fixed-probability *erasures* applied to the symbols of σ, x, and y. Using hypercontractivity bounds for the binary erasure channel [11], we conclude that ϕ fails the test on a significantly smaller fraction of queries than f does.

In the special case when the target group is $\mathbb{Z}_2$, the soundness error of D_f can be directly shown to be within a constant factor of the soundness error of the Dinur-Golubev tester (even though the two tests are different). The main motivating applications for function partitionability, however, concern real-valued functions [6]. The analysis of our $\oplus$-partionability testers for such functions relies on Theorem 1.

The idea of soundness analysis by iterative plurality decoding was introduced by Ben-Sasson et al. [2] and used by Shpilka and Wigderson [14] in the context of randomness-efficient linearity testing.

Testing partitionability

The main ingredient in our $\oplus$-partitionability algorithms is the direct 2-sum test D_f. The structure of this test allows us to efficiently detect a pair of variables x_s, x_t that must fall in the same component of the partition in any far from $\oplus$-partitionable function, effectively reducing the instance size by one variable.

Our $\otimes$-partitionability test looks for an input variable that is influential in at least two of the output coordinates of f. The analysis of this test is based on Lemma 24, which states that such a variable must exist in any far from partitionable function.

Organization

Section 2 outlines the proof of Theorem 1 in the case when the domain is the Boolean hypercube. The analysis is based on the convergence of the iterative decoder (Lemma 6), which is proved in Section 3. To prove Theorem 1 we use a reduction from testing functions over arbitrary product domains to testing functions on the hypercube (See the full version [5] for details of this reduction). Sections 4 and 5 describe and analyze the partitionability testers for direct sum and direct product, respectively.

Definitions and Notation

Let $\mathcal{D} \doteq \mathcal{D}_1 \times \ldots \times \mathcal{D}_n$ be a finite set. For strings $x, y \in \mathcal{D}$ and a set of indices $S \subseteq [n]$, let x_S to refer to the projection of x onto the coordinates in S. For strings $x^{(1)}, \ldots x^{(k)} \in \mathcal{D}$, and a partition $S_1, \ldots, S_k$ of $[n]$, let $x^{(1)}_{S_1} \ldots x^{(k)}_{S_k}$ be the string in $\mathcal{D}$ that is identical to $x^{(i)}$ on indices in S_i. For a bipartition $(S, \overline{S})$, we often write $x_S y$ instead of $x_S y_{\bar{S}}$.

In sections 2 and 3 we identify a bipartition $(S, \overline{S})$ of $[n]$ with its indicator vector $\sigma \in \{0,1\}^n$, and write $D_f(\sigma; x, y)$ instead of $D_f(S, \overline{S}; x, y)$, and x_σ instead of x_S.

We extend the definition of D_f to pairs of disjoint sets (S, T) that do not necessarily partition $[n]$ as

$$D_f(S, T; x, y) \doteq f(x) - f(y_S x) - f(y_T x) + f(y_{S \cup T} x).$$

2 Direct Sum Test for Functions on the Boolean Hypercube

The following dual characterization of direct sums motivates our test.

▶ **Fact 4.** *A function $f : \{0,1\}^n \to \mathcal{G}$ is a direct sum if and only if $D_f(\pi; x, y) = 0$ for every choice of $x, y, \pi \in \{0,1\}^n$.*

Proof of Fact 4. The "only if" direction is immediate from the definition of a direct sum. We prove the "if" direction. Let f be such that $D_f(\pi; x, y) = 0$ for every choice of $x, y, \pi \in \{0,1\}^n$. Fix $y \in \{0,1\}^n$. For every $x \in \{0,1\}^n$ we can write $f(x)$ as

$$\begin{aligned} f(x) &= f(x_{\{1\}}y) + f(y_{\{1\}}x) - f(y) \\ &= f(x_{\{1\}}y) + f(x_{\{2\}}y) + f(y_{\{1,2\}}x) - 2f(y) \\ &\quad \vdots \\ &= f(x_{\{1\}}y) + f(x_{\{2\}}y) + \ldots + f(x_{\{n\}}y) - (n-1)f(y). \end{aligned}$$

Therefore, f is a direct sum. ◀

Algorithm 1 Direct sum test for functions over $\{0,1\}^n$.

```
Oracle : $f\colon \{0,1\}^n \to \mathcal{G}$
Sample $x, y, \pi \in \{0,1\}^n$ independently and uniformly at random.
If $f(x) + f(y) - f(x_\pi y) - f(y_\pi x) = 0$, accept.
Else, reject.
```

By Fact 4, the test accepts every direct sum with probability 1. The following proposition establishes soundness of the test. Let $\rho(f)$ denote the probability that Algorithm 1 rejects the function f. That is, $\rho(f) \doteq \Pr_{x,y,\pi}[D_f(\pi; x, y) \neq 0]$.

▶ **Proposition 5** (Soundness). *There exist a universal constant $\eta \in [0,1]$ such that for every function $f : \{0,1\}^n \to \mathcal{G}$,*

$$\rho(f) \geq \min(\delta/4, \eta),$$

where δ is the distance between f and the set of direct sums.

▶ **Lemma 6** (Iterative decoding). *There exists a universal constant $\eta \in [0,1]$ such that for every function $f : \{0,1\}^n \to \mathcal{G}$ with $\rho(f) < \eta$, there exists a function $\phi : \{0,1\}^n \to \mathcal{G}$ such that:*

(i) *the function ϕ is $2\rho(f)$-close to f, and*
(ii) $\rho(\phi) \leq \rho(f)/2$.

Proof of Proposition 5. Iteratively applying Lemma 6 results in a sequence of functions $f = f_0, f_1, \ldots$, such that for all $t \geq 1$, (i) the distance between f_t and f_{t-1} is at most $2\rho(f_{t-1})$, and (ii) $\rho(f_t) \leq \rho(f_{t-1})/2$. The probability that the test rejects a function is a discrete quantity. So, by (ii), there must exist an integer t such that $\rho(f_t) = 0$. That is, $D_{f_t}(\pi; x, y) = 0$ for every choice of $x, y, \pi \in \{0,1\}^n$. By Fact 4 this means f_t is a direct sum. The distance between f and the direct sum f_t at most

$$\sum_{i=0}^{t-1} 2\rho(f_i) \leq 2\sum_{i=0}^{t-1} \rho(f)/2^i \leq 4\rho(f).$$ ◀

3 Analysis of Iterative Decoding

We begin with a sketch of the proof of Lemma 6. As mentioned in the introduction, the proof follows in the footsteps of the analysis of the BLR linearity test. We define $\phi(x)$ to be plurality$_{y,\pi}$ $f(x_\pi y) + f(y_\pi x) - f(y)$. Markov's inequality allows us to bound the distance between ϕ and f by $2\rho(f)$.

To show that Test 1 rejects ϕ with probability at most $\rho(f)/2$, we first show that for all but a $o(\rho(f))$-fraction of choices of x, $\phi(x)$ is defined by a strict majority that makes up at least 6/7-th of the plurality vote (See Claim 10). The fraction of x's that contribute to the plurality is at least the probability of a collision, i.e., $\Pr_{y,z,\sigma,\pi}[f(x_\pi y) + f(y_\pi x) - f(y) = f(x_\sigma z) + f(z_\sigma x) - f(z)]$. Using the algebraic identity in Claim 8, we can express this probability as

$$\Pr_{y,z,\pi,\sigma}[D_f(\pi; x_\pi z_{\overline{\pi}}, y) - D_f(\pi; y_\pi x_{\overline{\pi}}, z) + D_f(\pi \oplus \sigma; x_\sigma z_{\overline{\sigma}}, z_\sigma x_{\overline{\sigma}}) = 0].$$

The analysis of the BLR test also uses an analogous algebraic identity to bound the collision probability. The difference is that the resulting expression in the BLR analysis is made up of evaluations of the BLR test at points independent of x. This allows one to argue that the plurality vote is made up of a strict majority at all points x. In our setting, the arguments of D_f in the expression above are correlated with x. However, we can view these arguments as the result of passing x through a noisy binary erasure channel. This allows for the application of the hypercontractive inequality to bound the fraction of x for which the collision probability is less than 6/7.

We then show that for all but $o(\rho(f))$ choices of $x, y, \pi \in \{0,1\}^n$ there exist $z, w, \sigma \in \{0,1\}^n$ such that, (A) the value of $D_\phi(\pi; x, y) = \phi(x) - \phi(x_\pi y) - \phi(y_\pi x) + \phi(y)$ does not change after the following substitutions, and (B) the resulting expression post substitution evaluates to zero.

$$\begin{aligned}
\phi(x) &\leftarrow f(x_\sigma z) + f(z_\sigma x) - f(z)\\
\phi(x_\pi y) &\leftarrow f((x_\pi y)_\sigma(z_\pi w)) + f((z_\pi w)_\sigma(x_\pi y)) - f(z_\pi w)\\
\phi(y_\pi x) &\leftarrow f((y_\pi x)_\sigma(w_\pi z)) + f((w_\pi z)_\sigma(y_\pi x)) - f(w_\pi z)\\
\phi(y) &\leftarrow f(y_\sigma w) + f(w_\sigma y) - f(y).
\end{aligned} \tag{2}$$

It follows that the probability that ϕ is rejected by the test is $o(\rho(f))$.

By Claim 10, for all but $o(\rho(f))$ choices of x, y, π the substitutions do not change the value of $D_\phi(\pi; x, y)$ with probability at least 4/7. For (B), we use the algebraic identity in Claim 9 to rewrite the expression after substitution as

$$D_f(\pi \oplus \sigma; x_\sigma z, y_\sigma w) - D_f(\pi \oplus \overline{\sigma}; x_{\overline{\sigma}} z, y_{\overline{\sigma}} w) + D_f(\pi; z, w).$$

Again we show that the arguments of the D_f terms can be viewed as the result of passing x, y and π through independent binary erasure channels. Using the hypercontractive inequality, we conclude that for all but $o(\rho(f))$ choices of x, y and π the expression after substitution evaluates to zero for most choices of z, w and σ. By a union bound we can ensure that (A) and (B) hold simultaneously for the same z, w and σ.

The following technical lemma establishes the bounds we prove using the hypercontractivity of the binary erasure channel. The proof is presented in Section 3.1.

Let QUERIES$(\pi; x, y)$ denote the vector in $(\{0,1\}^n)^4$ whose entries are the four queries that Algorithm 1 makes when π, x, y is sampled. That is, QUERIES$(\pi; x, y) = (x, x_\pi y, y_\pi x, y)$.

▶ **Lemma 7.** *Let* $\textsc{Bad} \subset (\{0,1\}^n)^4$ *be a set such that the probability that* $\textsc{queries}(\pi, x, y)$ *lands in* $\textsc{Bad}$*, when* π, x, y *are chosen independently and uniformly at random, is* ρ*.*

(i) $\mu_x(A_1) \leq 21^2 \rho^{4/3}$*, where*

$$A_1 = \{x \mid \Pr_{\pi,y,z}[\textsc{queries}(\pi; x_\pi z_{\overline{\pi}}, y) \in \textsc{Bad}] \geq 1/21\}.$$

(ii) $\mu_x(A_2) \leq 21^2 \rho^{4/3}$*, where*

$$A_2 = \{x \mid \Pr_{\pi,\sigma,z}[\textsc{queries}(\pi \oplus \sigma; x_\sigma z_{\overline{\sigma}}, z_\sigma x_{\overline{\sigma}}) \in \textsc{Bad}] \geq 1/21\}.$$

(iii) $\mu_{\pi,x,y}(A_3) \leq 7^2 \rho^{2/(1+\sqrt{2/3})} \leq 7^2 \rho^{1.1}$*, where*

$$A_3 = \{(\pi, x, y) \mid \Pr_{\sigma,z,w}[\textsc{queries}(\pi \oplus \sigma; x_\sigma z, y_\sigma w) \in \textsc{Bad}] \geq 1/7\}.$$

(iv) $\mu_\pi(A_4) \leq 7^2 \rho^{4/3}$*, where*

$$A_4 = \{\pi \mid \Pr_{z,w}[\textsc{queries}(\pi; z, w) \in \textsc{Bad}] \geq 1/7\}.$$

We will also need the following algebraic identities.

▷ Claim 8. The following identity holds:

$$D_f(\pi; x, y) - D_f(\sigma; x, z) = D_f(\pi; x_\pi z_{\overline{\pi}}, y) - D_f(\pi; y_\pi x_{\overline{\pi}}, z) + D_f(\pi \oplus \sigma; x_\sigma z_{\overline{\sigma}}, z_\sigma x_{\overline{\sigma}}).$$

Proof of Claim 8. The claim follows by adding the following two identities:

$$\begin{aligned} D_f(\pi; x, y) - D_f(\pi; x, z) &= -f(x_\pi y_{\overline{\pi}}) - f(y_\pi x_{\overline{\pi}}) + f(y) + f(x_\pi z_{\overline{\pi}}) + f(z_\pi x_{\overline{\pi}}) - f(z) \\ &= f(x_\pi z_{\overline{\pi}}) + f(y) - f(x_\pi y_{\overline{\pi}}) - f(y_\pi z_{\overline{\pi}}) \\ &\quad - f(y_\pi x_{\overline{\pi}}) - f(z) + f(y_\pi z_{\overline{\pi}}) + f(z_\pi x_{\overline{\pi}}) \\ &= D_f(\pi; x_\pi z_{\overline{\pi}}, y) - D_f(\pi; y_\pi x_{\overline{\pi}}, z) \end{aligned}$$

$$\begin{aligned} D_f(\pi; x, z) - D_f(\sigma; x, z) &= -f(x_{\pi\sigma} x_{\pi\overline{\sigma}} z_{\overline{\pi}\sigma} z_{\overline{\pi\sigma}}) - f(z_{\pi\sigma} z_{\pi\overline{\sigma}} x_{\overline{\pi}\sigma} x_{\overline{\pi\sigma}}) + f(x) + f(z) \\ &\quad + f(x_{\pi\sigma} z_{\pi\overline{\sigma}} x_{\overline{\pi}\sigma} z_{\overline{\pi\sigma}}) + f(z_{\pi\sigma} x_{\pi\overline{\sigma}} z_{\overline{\pi}\sigma} x_{\overline{\pi\sigma}}) - f(x) - f(z) \\ &= D_f(\pi \oplus \sigma; x_\sigma z_{\overline{\sigma}}, z_\sigma x_{\overline{\sigma}}) \end{aligned}$$

◀

To analyze the substitutions (2) we set $D_{\phi,f}(\pi; x, y) = \phi(x) - f(x_\pi y) - f(y_\pi x) + f(x)$. In particular, $D_{f,f} = D_f$.

▷ Claim 9 (16-point identity). The following identity holds:

$$\begin{aligned} &D_{\phi,f}(\sigma; x, z) - D_{\phi,f}(\sigma; x_\pi y, w_\pi z) - D_{\phi,f}(\sigma; y_\pi x, z_\pi w) + D_{\phi,f}(\sigma; y, w) \\ &\qquad = D_\phi(\pi; x, y) - D_f(\pi \oplus \sigma; x_\sigma z, y_\sigma w) - D_f(\pi \oplus \overline{\sigma}; x_{\overline{\sigma}} z, y_{\overline{\sigma}} w) + D_f(\pi; z, w). \end{aligned}$$

Proof of Claim 9. We write $xyzw$ to denote the string $x_{\pi\sigma} y_{\pi\overline{\sigma}} z_{\overline{\pi}\sigma} w_{\overline{\pi\sigma}}$. With this notation,

$$\begin{array}{lllll} +D_{\phi,f}(\sigma; x, z) = & +\phi(xxxx) & -f(xzxz) & -f(zxzx) & +f(zzzz) \\ -D_{\phi,f}(\sigma; x_\pi y, w_\pi z) = & -\phi(xxyy) & +f(xwyz) & +f(wxzy) & -f(wwzz) \\ -D_{\phi,f}(\sigma; y_\pi x, z_\pi w) = & -\phi(yyxx) & +f(yzxw) & +f(zywx) & -f(zzww) \\ +D_{\phi,f}(\sigma; y, w) = & +\phi(yyyy) & -f(ywyw) & -f(wywy) & +f(wwww) \\ & \parallel & \parallel & \parallel & \parallel \\ & +D_\phi(\pi; x, y) & -D_f(\pi \oplus \sigma; x_\sigma z, y_\sigma w) & -D_f(\pi \oplus \overline{\sigma}; x_{\overline{\sigma}} z, y_{\overline{\sigma}} w) & +D_f(\pi; z, w) \end{array}$$

The identity states that the column sums and the row sums add up. ◁

Proof of Lemma 6. Let ϕ be a function defined as $\phi(x) = \text{plurality}_{y,\pi}\, f(x_\pi y) + f(y_\pi x) - f(y)$. That is, $\phi(x)$ is the most frequent value of $f(x_\pi y) + f(y_\pi x) - f(y)$, where $y, \pi \in \{0,1\}^n$. Ties are broken arbitrarily. We show that ϕ satisfies the hypothesis of the lemma.

(i) ϕ is $2\rho(f)$-close to f: For $x \in \{0,1\}^n$, let $\rho_x \doteq \Pr_{y,\pi}[f(x) \neq f(x_\pi y) + f(y_\pi x) - f(y)]$. Note that $\mathbb{E}_x[\rho_x] = \rho(f)$, and that if $\rho_x < 1/2$ then $f(x) = \phi(x)$. Thus, by Markov's inequality,

$$\Pr_x[f(x) \neq \phi(x)] \leq \Pr_x[\rho_x \geq 1/2] \leq 2\rho(f).$$

(ii) $\rho(\phi) \leq \rho(f)/2$: We begin by showing that with probability $\rho(f)/12$ over the choice of x, the plurality that defines $\phi(x)$ is a majority made up of 6/7-th of the votes. Then $\Pr_{\pi,y}[D_{\phi,f}(\pi; x, y) = 0]$ is the fraction of votes that constitute the plurality defining $\phi(x)$. Let

$$\textsc{Weak-Maj} = \left\{x \;\middle|\; \Pr_{y,\pi}[D_{\phi,f}(\pi; x, y) \neq 0] \geq 1/7\right\}.$$

▷ Claim 10 (Strong Majority). $\mu_x(\textsc{Weak-Maj}) \leq \rho(f)/12$.

Proof. The fraction of votes that contribute to the plurality $\Pr_{y,\pi}[D_{\phi,f}(\pi; x, y) = 0]$ is an upper bound on the collision probability $\Pr_{y,\pi,z,\sigma\in\{0,1\}^n}[D_{\phi,f}(\pi; x, y) = D_{\phi,f}(\sigma; x, z)]$. This is because

$$\begin{aligned}\Pr_{y,\pi,z,\sigma}[D_{\phi,f}(\pi; x, y) = D_{\phi,f}(\sigma; x, z)] &= \sum_{\gamma\in\mathcal{G}} \Pr_{y,\pi}[D_{\phi,f}(\pi; x, y) = \gamma]^2 \\ &\leq \max_{\gamma\in\mathcal{G}} \Pr_{y,\pi}[D_{\phi,f}(\pi; x, y) = \gamma] \\ &= \Pr_{y,\pi}[D_{\phi,f}(\pi; x, y) = 0].\end{aligned}$$

The final equality holds because $\phi(x) = \arg\max_{\beta\in\mathcal{G}} \Pr_{y,\pi}[\beta - f(x_\pi y) - f(y_\pi x) + f(y) = 0]$. We showed that

$$\mu_x(\textsc{Weak-Maj}) \leq \mu_x\{x \mid \Pr_{y,\pi,z,\sigma}[D_{\phi,f}(\pi; x, y) \neq D_{\phi,f}(\sigma; x, z)] \geq 1/7\}.$$

We now use Lemma 7 to bound the right hand side. Let $\textsc{Bad}_f \subset (\{0,1\}^n)^4$ be the set of queries on which D_f fails, namely

$$\textsc{Bad}_f = \{\textsc{queries}(\pi; x, y) \mid D_f(\pi; x, y) \neq 0\}.$$

This is a set of measure $\mu_{\pi,x,y}(\textsc{Bad}) = \rho(f)$. Since $D_{\phi,f}(\pi; x, y) - D_{\phi,f}(\sigma; x, z) = D_f(\pi; x, y) - D_f(\sigma; x, z)$, by the algebraic identity in Claim 8 and a union bound, we have

$$\begin{aligned}\mu_x(\textsc{Weak-Maj}) &\leq \mu_x\{x \mid \Pr_{\pi,\sigma,y,z}[D_{\phi,f}(\pi; x, y) - D_{\phi,f}(\sigma; x, z) \neq 0] \geq 1/7\} \\ &\leq \mu_x\{x \mid \Pr_{\pi,y,z}[D_f(\pi; x_\pi z_{\overline{\pi}}, y) \neq 0] \geq 1/21\} \\ &\qquad + \;\mu_x\{x \mid \Pr_{\pi,y,z}[D_f(\pi; y_\pi x_{\overline{\pi}}, z) \neq 0] \geq 1/21\} \\ &\qquad + \;\mu_x\{x \mid \Pr_{\pi,\sigma,z}[D_f(\pi \oplus \sigma; x_\sigma z_{\overline{\sigma}}, z_\sigma x_{\overline{\sigma}}) \neq 0] \geq 1/21\} \\ &= \mu_x(A_1) + \mu_x(A_1) + \mu_x(A_2),\end{aligned}$$

where A_1 and A_2 are the sets

$$\begin{aligned}A_1 &= \{x \mid \Pr_{\pi,y,z}[\textsc{queries}(\pi; x_\pi z_{\overline{\pi}}, y) \in \textsc{Bad}_f] \geq 1/21\}, \\ A_2 &= \{x \mid \Pr_{\pi,\sigma,z}[\textsc{queries}(\pi \oplus \sigma; x_\sigma z_{\overline{\sigma}}, z_\sigma x_{\overline{\sigma}}) \in \textsc{Bad}_f] \geq 1/21\}.\end{aligned}$$

By Lemma 7 we get that $\mu_x(\textsc{Weak-Maj}) \leq \rho(f)/12$, for small enough η. ◁

We are now ready to prove that $\rho(\phi) = \Pr_{\pi,x,y}[D_\phi(\pi; x, y) \neq 0] \leq \rho(f)/2$. In order to do so we define a set $\textsc{Bad}_\phi$ of triples (π, x, y) such that $\mu_{\pi,x,y}(\textsc{Bad}_\phi) \leq \rho(f)/2$, and if $(\pi, x, y) \notin \textsc{Bad}_\phi$ then $D_\phi(\pi, x, y) = 0$.

Let A_3 and A_4 denote the sets

$$A_3 = \{(\pi, x, y) \mid \Pr_{\sigma,z,w}[\textsc{queries}(\pi \oplus \sigma; x_\sigma z, y_\sigma w) \in \textsc{Bad}_f] \geq 1/7\},$$
$$A_4 = \{\pi \mid \Pr_{z,w}[\textsc{queries}(\pi; z, w) \in \textsc{Bad}_f] \geq 1/7\}.$$

Let $\textsc{Bad}_\phi$ be the set

$$\{(\pi, x, y) \mid (\text{One of } x, y, x_\pi y, y_\pi x \text{ lies in } \textsc{Weak-Maj}) \text{ or } ((\pi, x, y) \in A_3) \text{ or } (\pi \in A_4)\}\,.$$

By Lemma 7, $\mu_{\pi,x,y}(A_3) \leq \rho(f)/12$, and $\mu_\pi(A_4) \leq \rho(f)/12$, for small enough η. As $x, y, x_\pi y, y_\pi x$ are all random, by a union bound we have

$$\mu_{\pi,x,y}(\textsc{Bad}_\phi) \leq 4\mu_x(\textsc{Weak-Maj}) + \mu_{\pi,x,y}(A_3) + \mu_\pi(A_4) \leq \rho(f)/2.$$

All that remains to show is that if $(\pi, x, y) \notin \textsc{Bad}_\phi$, $D_\phi(\pi; x, y) = 0$. On rearranging the terms in the algebraic identity of Claim 9, we get

$$\begin{aligned} D_\phi(\pi; x, y) = {} & D_{\phi,f}(\sigma; x, z) - D_{\phi,f}(\sigma; x_\pi y, w_\pi z) - D_{\phi,f}(\sigma; y_\pi x; z_\pi w) + D_{\phi,f}(\sigma; y, w) \\ & + D_f(\pi \oplus \sigma; x_\sigma z, y_\sigma w) + D_f(\pi \oplus \overline{\sigma}; x_{\overline{\sigma}} z, y_{\overline{\sigma}} w) - D_f(\pi; z, w). \end{aligned} \tag{3}$$

Fix a triple $(\pi, x, y) \notin \textsc{Bad}_\phi$. We show that $D_\phi(\pi, x, y) = 0$, by showing that there exists a choice of σ, z and w for which the right hand side of Equation (3) evaluates to zero. By Equation (3) and a union bound,

$$\begin{aligned} \Pr_{\sigma,z,w}[D_\phi(\pi; x, y) \neq 0] &= \Pr_{\sigma,z,w}\left[\begin{array}{l} \quad D_{\phi,f}(\sigma; x, z) \neq 0 \\ \text{or } D_{\phi,f}(\sigma; x_\pi y, w_\pi z) \neq 0 \\ \text{or } D_{\phi,f}(\sigma; y_\pi x; z_\pi w) \neq 0 \\ \text{or } D_{\phi,f}(\sigma; y, w) \neq 0 \\ \text{or } D_f(\pi \oplus \sigma; x_\sigma z, y_\sigma w) \neq 0 \\ \text{or } D_f(\pi \oplus \overline{\sigma}; x_{\overline{\sigma}} z, y_{\overline{\sigma}} w) \neq 0 \\ \text{or } D_f(\pi; z, w) \neq 0 \end{array}\right] \\ &< 4/7 + \Pr_{\sigma,z,w}\left[\begin{array}{l} \quad \textsc{queries}(\pi \oplus \sigma; x_\sigma z, y_\sigma w) \in \textsc{Bad}_f \\ \text{or } \textsc{queries}(\pi \oplus \overline{\sigma}; x_{\overline{\sigma}} z, y_{\overline{\sigma}} w) \in \textsc{Bad}_f \\ \text{or } \textsc{queries}(\pi; z, w) \in \textsc{Bad}_f \end{array}\right] \\ &< 4/7 + 3/7 = 1. \end{aligned}$$

The first inequality holds because $x, x_\pi y, y_\pi x, y \notin \textsc{Weak-Maj}$, and the second inequality holds because $(\pi, x, y) \notin A_3$ and $\pi \notin A_4$. Since the probability $\Pr_{\sigma,z,w}[D_\phi(\pi; x, y) \neq 0]$ is either 0 or 1, it must be that $D_\phi(\pi; x, y) = 0$. Therefore,

$$\rho(\phi) = \Pr_{\pi,x,y}[D_\phi(\pi; x, y) \neq 0] \leq \mu_{\pi,x,y}(\textsc{Bad}_\phi) \leq \rho(f)/2. \qquad \blacktriangleleft$$

3.1 Proof of Lemma 7

We begin with some preliminaries on discrete channels and hypercontractivity. For a motivating discussion on hypercontractivity and a proof of Fact 14 below see Chapter 9 of [12].

▶ **Definition 11** (Discrete channels). *A* discrete channel *is a triple* $(\mathcal{U}, P, \mathcal{V})$*, where* $\mathcal{U}$ *and* $\mathcal{V}$ *are finite sets representing the* input alphabet *and* output alphabet*, and* P *is a* $\mathcal{U} \times \mathcal{V}$ *probability* transition matrix *that describes the distribution of the output conditioned on the input. The* composition *of two channels* $(\mathcal{U}, P_1, \mathcal{V})$ *and* $(\mathcal{V}, P_2, \mathcal{W})$ *is the channel* $(\mathcal{U}, P_1 \cdot P_2, \mathcal{W})$*, where* $\cdot$ *is matrix multiplication.*

The binary erasure channel will play an important role in the proof of Lemma 7.

▶ **Definition 12** (Binary Erasure Channel). *The* binary erasure channel $BEC(e)$ *with erasure probability* e *has input alphabet* $\{0,1\}$ *and output alphabet* $\{0,1,\perp\}$*, and probability transition matrix* $P(x|x) = 1-e, P(\perp|x) = e$*.*

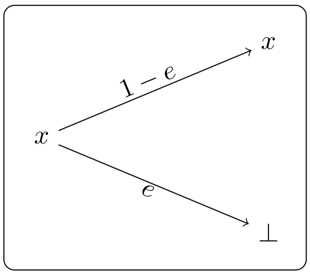

Figure 1 The binary erasure channel $BEC(e)$.

For a real valued random variable U and $p \geq 1$, we denote the p-norm of U by $\|U\|_p \doteq \mathbb{E}_U[|U|^p]^{1/p}$.

▶ **Definition 13** (Hypercontractivity). *For* $1 \leq q \leq p$*, A pair of random variables* (U,V) *is* (p,q)*-hypercontractive if for every pair of real valued functions* f, g*,*

$$\mathbb{E}[f(U)g(V)] \leq \|f(U)\|_{p'} \|\|g(V)\|_q,$$

where $p' = p/(p-1)$ *is the Hölder conjugate of* p*.*

▶ **Fact 14** (Tensorisation [7]). *If* (U_1, V_1) *and* (U_2, V_2) *are independent random variables that are* (p,q)*-hypercontractive, then* $((U_1,U_2),(V_1,V_2))$ *is* (p,q) *hypercontractive.*

▶ **Theorem 15** (Hypercontractivity of $BEC(e)$ [11]). *Let* U *be distributed uniformly over* $\{0,1\}$ *and let* $V \in \{0,1,\perp\}$ *denote the output of* $BEC(e)$ *on input* U*. Then* (U,V) *is* (p,q)*-hypercontractive for all* $1 \leq q \leq p$ *such that*

$$\frac{q-1}{p-1} \geq 1-e.$$

▶ **Fact 16** (Composition). *Let* $(\mathcal{U}, P_1, \mathcal{V})$ *and* $(\mathcal{V}, P_2, \mathcal{W})$ *be two channels. Let* U *be a random variable over* $\mathcal{U}$*. Let* V *be the random variable that represents the output of the first channel on input* U*, and* W *the random variable that represents the output of the second channel on input* V*. If* (U,V) *is* (p,q)*-hypercontractive then so is* (U,W)*.*

Proof of Fact 16. Let $f : \mathcal{U} \to \mathbb{R}$ and $g : \mathcal{W} \to \mathbb{R}$ be arbitrary functions. Since $U \to V \to W$ is a markov chain, we have

$$\mathbb{E}_{U,W}[f(U)g(W)] = \mathbb{E}_{U,V}[f(U)E_W[g(W) \mid V]] \leq \|f(U)\|_{p'} \| \mathbb{E}_W[g(W) \mid V]\|_q,$$

where the inequality holds because (U,V) is (p,q) hypercontractive.

Now, by Jensen's inequality,

$$\mathbb{E}_V[\mathbb{E}_W[g(W) \mid V]^q]^{1/q} \le \mathbb{E}_V[\mathbb{E}_W[g(W)^q \mid V]]^{1/q} = \mathbb{E}_W[g(W)^q]^{1/q} = \|g(W)\|_q$$

Therefore, (U, W) is (p, q) hypercontractive. ◀

The following claim captures the small-set expansion interpretation of hypercontractivity [1] in the form used in the proof of Lemma 7.

▷ Claim 17. Let U, V be random variables that take values in $\mathcal{U}$ and $\mathcal{V}$ respectively. Let $B \subset \mathcal{V}$ be a set such that $\Pr[V \in B] = \rho$. Let $A \subset \mathcal{U}$ denote the set $\{u \mid \Pr[V \in B \mid U = u] \ge \theta\}$. If (U, V) is (p, q) hypercontractive, then $\Pr[U \in A] \le \rho^{p/q}/\theta^p$.

Proof. Let 1_A and 1_B denote the indicator functions of the sets A and B. Since (U, V) are (p, q) hypercontractive,

$$\theta \cdot \Pr[U \in A] \le \Pr[V \in B \mid U \in A]\Pr[U \in A] = \mathbb{E}[1_A(U)1_B(V)] \le \|1_A(U)\|_{p'}\|1_B(V)\|_q,$$

where $p' = p/(p-1)$. Note that $\|1_B(V)\| = \rho^{1/q}$, and $\|1_A(U)\|_{p'} = \Pr[U \in A]^{1/p'}$. Therefore, $\Pr[U \in A]^{1/p} \le \rho^{1/q}/\theta$, that is, $\Pr[U \in A] \le \rho^{p/q}/\theta^p$. ◁

Proof of Lemma 7. We need to bound the probabilities of four sets of the form

$$\{u \in \Sigma^n \mid \Pr[\textsc{queries}(\psi(u)) \in \textsc{bad} \mid U = u]\} \ge \theta,$$

where ψ is some (randomized) function. All bounds of the form $\theta^{-2}\rho^{2/q}$ will follow from Claim 17 by showing that the channel $U \to \textsc{queries}(\psi(u))$ is $(2, q)$-hypercontractive for a suitable choice of q ($q = 3/2$ for parts (i), (ii), (iv) and $q = 1 + \sqrt{2/3}$ for part (iii)).

The *channel* $(\Sigma^n, P_n, \{0,1\}^{n\times 4})$ that maps $u \in \Sigma^n$ to $\textsc{queries}(\psi(u)) \in \{0,1\}^{n\times 4}$ acts independently on the symbols $u_1, \ldots, u_n$. In all cases, the i-th bits of the four queries (q_1, q_2, q_3, q_4) are obtained by applying the one-dimensional channel P_1 to u_i. Therefore, P_n tensorizes as $P_n = P_1^{\otimes n}$. By Fact 14, it is sufficient to show that the channel P_1 is hypercontractive. We may and will therefore assume, without loss of generality, that $n = 1$.

We now demonstrate how each of the four channels of interest can be decomposed into a binary erasure channel with constant erasure probability ($e = 1/2$ in parts (i), (ii), (iv) and $e = 1 - \sqrt{2/3}$ in part (iii)) and some other fixed channel. The Lemma then follows from Fact 16 and Theorem 15 with $q = 2 - e$.

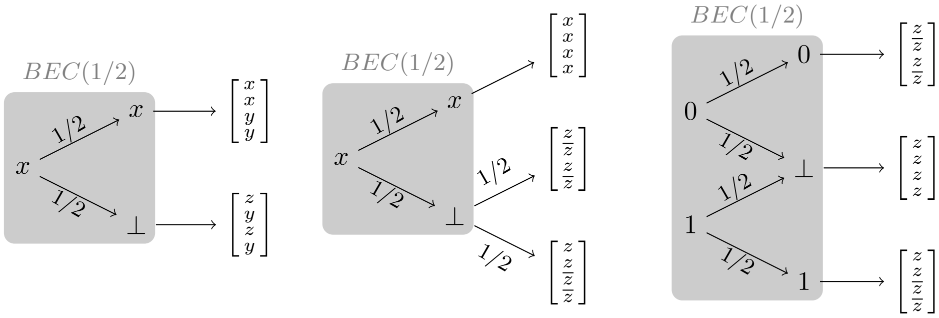

Figure 2 Channels (i) $x \to \textsc{queries}(\pi, x_\pi z_{\overline{\pi}}, y)$; (ii) $x \to \textsc{queries}(\pi \oplus \sigma, x_\sigma z_{\overline{\sigma}}, z_\sigma x_{\overline{\sigma}})$; (iv) $\pi \to \textsc{queries}(\pi, z, w)$. y and z are random bits.

(i) The channel $x \to \text{QUERIES}(\pi; x_\pi z_{\overline{\pi}}, y) = (x_\pi z_{\overline{\pi}}, x_\pi y_{\overline{\pi}}, y_\pi z_{\overline{\pi}}, y)$ from $\Sigma = \{0,1\}$ to $\{0,1\}^4$ can be decomposed in the following way: On input x the channel samples a random bit π and outputs $xxyy$ if $\pi = 1$, and $zyzy$ if $\pi = 0$ for random y and z. This channel can be alternatively described as $BEC(1/2)$ composed with a second channel that outputs $xxyy$ if there is no erasure and the independent symbol $zyzy$ otherwise. See Figure 2 (i).

(ii) The channel from $\Sigma = \{0,1\}$ to $\{0,1\}^4$ is of the form

$$x \to \text{QUERIES}(\pi \oplus \sigma, x_\sigma z_{\overline{\sigma}}, z_\sigma x_{\overline{\sigma}}) = \begin{pmatrix} x_{\pi\sigma} z_{\pi\overline{\sigma}} x_{\overline{\pi}\sigma} z_{\overline{\pi}\overline{\sigma}} \\ z_{\pi\sigma} z_{\pi\overline{\sigma}} x_{\overline{\pi}\sigma} x_{\overline{\pi}\overline{\sigma}} \\ x_{\pi\sigma} x_{\pi\overline{\sigma}} z_{\overline{\pi}\sigma} z_{\overline{\pi}\overline{\sigma}} \\ z_{\pi\sigma} x_{\pi\overline{\sigma}} z_{\overline{\pi}\sigma} x_{\overline{\pi}\overline{\sigma}} \end{pmatrix} = \begin{cases} xzxz, & \text{if } \pi\sigma = 1, \\ zzxx, & \text{if } \pi\overline{\sigma} = 1, \\ xxzz, & \text{if } \overline{\pi}\sigma = 1, \\ zxzx, & \text{if } \overline{\pi}\overline{\sigma} = 1, \end{cases}$$

where π, σ, z are random bits. We can alternatively describe it like this: If $z = x$, then output $xxxx$. If $z \neq x$ and $\pi \oplus \sigma = 1$, then output $zz\overline{z}\overline{z}$. If $z \neq x$ and $\pi \oplus \sigma = 0$, then output $zzzz$.

This channel can be factored through $BEC(1/2)$ as in Figure 2 (ii). If there is no erasure, the second channel outputs $xxxx$. If there is an erasure, then the second channel outputs $zz\overline{z}\overline{z}$ with probability $1/2$ and $z\overline{z}z\overline{z}$ with probability $1/2$.

(iii) The channel from $\Sigma = \{0,1\}^3$ to $\{0,1\}^4$ is of the form

$$\begin{pmatrix} \pi \\ x \\ y \end{pmatrix} \to \text{QUERIES}(\pi \oplus \sigma; x_\sigma z, y_\sigma w) = \begin{pmatrix} x_{\pi\sigma} z_{\pi\overline{\sigma}} x_{\overline{\pi}\sigma} z_{\overline{\pi}\overline{\sigma}} \\ y_{\pi\sigma} z_{\pi\overline{\sigma}} x_{\overline{\pi}\sigma} w_{\overline{\pi}\overline{\sigma}} \\ x_{\pi\sigma} w_{\pi\overline{\sigma}} y_{\overline{\pi}\sigma} z_{\overline{\pi}\overline{\sigma}} \\ y_{\pi\sigma} w_{\pi\overline{\sigma}} y_{\overline{\pi}\sigma} w_{\overline{\pi}\overline{\sigma}} \end{pmatrix} = \begin{cases} xyxy, & \text{if } \pi\sigma = 1, \\ zzww, & \text{if } \pi\overline{\sigma} = 1, \\ xxyy, & \text{if } \overline{\pi}\sigma = 1, \\ zwzw, & \text{if } \overline{\pi}\overline{\sigma} = 1. \end{cases}$$

Consider the composition of the following two channels. The first channel views the symbol πxy as three bits and independently applies $BEC(1/4)$ to π and $BEC(1 - \sqrt{2/3})$ to x and y. The second channel is described in Figure 3.

- If π is erased, the second channel outputs $zzzz$, for a uniform bit z. This corresponds to the event $z = w$ and $\sigma = 0$.
- If π is not erased but one of x, y is erased, the second channel samples a uniform bit $z \in \{0,1\}$ and outputs $z\overline{z}z\overline{z}$ if $\pi = 0$ and $zz\overline{z}\overline{z}$ if $\pi = 1$. This corresponds to the event $z \neq w$ and $\sigma = 0$.
- If there are no erasures, then the second channel outputs $xyxy$ if $\pi = 1$, and $xxyy$ if $\pi = 0$. This corresponds to the event $\sigma = 1$.

The first channel is $BEC(\frac{1}{4}) \otimes BEC(1 - \sqrt{2/3}) \otimes BEC(1 - \sqrt{2/3})$. Since $1 - \sqrt{2/3} \leq 1/4$, by Fact 14 it inherits the hypercontractivity parameters of $BEC(1 - \sqrt{2/3})$.

(iv) The channel $\pi \to \text{QUERIES}(\pi, z, w)$ from $\Sigma = \{0,1\}$ to $\{0,1\}^4$ outputs $zzww$ if $\pi = 1$ and $zwzw$ if $\pi = 0$ for random bits z and w. Alternatively, the channel can be described as a uniform choice between $zzzz$ and $zz\overline{z}\overline{z}$ when $\pi = 1$ and a uniform choice between $zzzz$ and $z\overline{z}z\overline{z}$ when $\pi = 0$. This can be modeled as the composition of $BEC(1/2)$ and a second channel that outputs $zzzz$ if there is an erasure, and either $zz\overline{z}\overline{z}$ or $z\overline{z}z\overline{z}$ depending on the value of π_i otherwise. See Figure 2 (iv). ◀

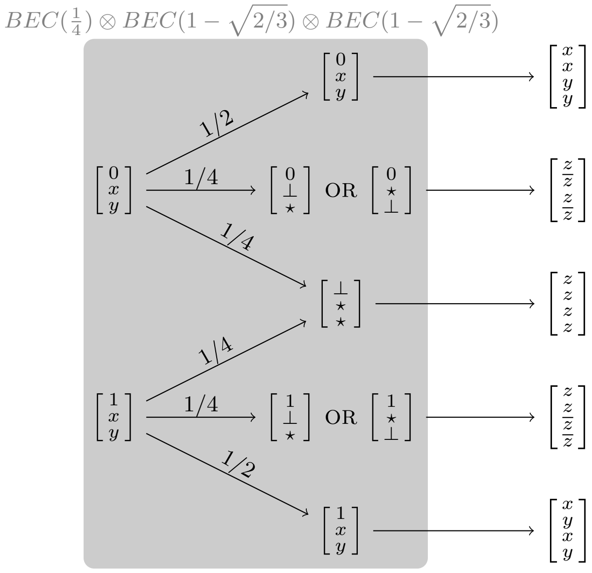

Figure 3 (iii) Channel $(\pi, x, y) \to \text{QUERIES}(\pi \oplus \sigma, x_\sigma z, y_\sigma w)$. A $\star$ represents any of $\{0, 1, \perp\}$ and z is a random bit.

4 Testing $\oplus$-Partitionability

Recall that a function $f : \mathcal{D} = \mathcal{D}_1 \times \cdots \times \mathcal{D}_n \to \mathcal{G}$ is *k-$\oplus$-partitionable* if there exists a k-partition $S_1, \ldots, S_k$ of $[n]$, and functions $f_1, \ldots, f_k$ such that $f(x) = f_1(x_{S_1}) + \ldots + f_k(x_{S_k})$, for all $x \in \mathcal{D}$.

Recall that for disjoint sets $S, T \subseteq [n]$,

$$D_f(S, T; x, y) \doteq f(x) - f(y_S x) - f(y_T x) + f(y_{S \cup T} x)$$

The following claim is an immediate consequence of Theorem 1, and allows us to determine whether a function f is $\oplus$-partitionable with respect to a fixed partition $S_1, \ldots, S_k$.

▷ Claim 18. Let $(S, \overline{S})$ be a random coarsening of a k-partition $(S_1, \ldots, S_k)$ obtained by adding each S_i to S with probability $1/2$. If f is ϵ-far from $\oplus$-partitionable with respect to $S_1, \ldots, S_k$, then $D_f(S, \overline{S}; x, y)$ is nonzero with probability $\Omega(\epsilon)$.

To determine whether a function is k-$\oplus$-partitionable, our testers use the 4-query test D_f to group together variables that cannot occur in different partition components. If the tester finds fewer than k groups, it rejects, otherwise it accepts.

4.1 Adaptive Test for $\oplus$-Partitionability

Our test for k-$\oplus$-partitionability (Algorithm 3) seeks to identify a pair of *contractable* variables s, t that must fall in the same component of a partition. Variables s and t are then contracted and the test is repeated until either fewer than k variables are left (giving a certificate of non-partionability) or no contractable candidates can be found.

A sufficient condition for contractability is that $D_f(\{s\}, \{t\}; x, y)$ is nonzero for some assignment x, y. We start by splitting the variables into k components $S_1, \ldots, S_k$ arbitrarily and zero-testing $D_f(S, \bar{S}; x, y)$ for a random coarsening of the components into $S, \bar{S}$. By Claim 18, the zero-test fails with probability at least $\Omega(\epsilon)$, where ϵ is the distance between f and the set of functions that are $\oplus$-partitionable with respect to $S_1, \ldots, S_k$.

Once such a bipartition $S, \bar{S}$ is identified, s and t can be identified via binary search using Algorithm 2 below. The same idea was used by Blais [3] to identify an influential variable in his junta test. Our t is in fact the influential variable in the function $g(x_{[n]\setminus S}) = f(x) - f(y_S x)$, for fixed x_S, y_S, returned by Blais' test.

Algorithm 2 Violating pair adaptive search.

```
Oracle : $f\colon \mathcal{D}_1 \times \cdots \times \mathcal{D}_n \to \mathcal{G}$
Input  : $(S, T; x, y)$ such that $D_f(S, T; x, y) \neq 0$.
Output : $(\{s\}, \{t\})$ such that $D_f(\{s\}, \{t\}; x', y') \neq 0$ for some $x', y'$.
If $|S| = |T| = 1$, output $S, T$.
If $|T| = 1$, swap $S$ and $T$.
do
    Split $T$ into two subsets $T'$ and $T''$ of (almost) equal size.
    If $D_f(S, T'; x, y) \neq 0$, recursively run on input $(S, T'; x, y)$.
    Otherwise, recursively run on input $(S, T''; y, x)$.
```

The correctness of Algorithm 2 is based on the following identity.

▷ Claim 19. $D_f(S, T \cup T'; x, y) = D_f(S, T; x, y) + D_f(S, T'; y, x)$ for disjoint sets S, T, T'.

Proof. Without loss of generality take $S = \{1\}$, $T = \{2\}$, $T' = \{3\}$ and assume there are no other inputs (they are all fixed). By the definition of D_f,

$$\begin{aligned} D_f(\{1\}, \{2\}; x, y) &= f(x_1x_2x_3) + f(y_1y_2x_3) - f(x_1y_2x_3) - f(y_1x_2x_3) \\ D_f(\{1\}, \{3\}; y, x) &= f(y_1y_2y_3) + f(x_1y_2x_3) - f(y_1y_2x_3) - f(x_1y_2y_3) \\ -D_f(\{1\}, \{2,3\}; x, y) &= -f(x_1x_2x_3) - f(y_1y_2y_3) + f(x_1y_2y_3) + f(y_1x_2x_3). \end{aligned}$$

The terms on the right hand side cancel out. ◁

▶ **Lemma 20.** *Algorithm 2 is correct and has query complexity at most* $4(\lceil \log |S| \rceil + \lceil \log |T| \rceil)$.

Proof. The correctness follows from Claim 19 and from the symmetry of D_f in the S, T inputs. As for the query complexity, the algorithm makes four queries (in fact at most two additional queries) in each iteration, and each iteration shrinks one of the original inputs S, T by half. ◀

In the following algorithm we let $\mathcal{P}(S_1, \ldots, S_k)$ be the distribution on disjoint pairs of sets $(S, \bar{S})$ from Claim 18.

Algorithm 3 Adaptive tester for k-$\oplus$-paritionability.

```
Oracle : $f\colon \mathcal{D}_1 \times \ldots \times \mathcal{D}_n \to \mathcal{G}$
Input  : Size $k$ of partition
If $f$ has fewer than $k$ variables, output "not partitionable".
Otherwise, partition variables arbitrarily into $k$ sets $S_1, \ldots, S_k$.
repeat
    Choose sets $(S, \bar{S})$ at random from $\mathcal{P}(S_1, \ldots, S_k)$.
    Choose random inputs $x, y$.
until $D_f(S, T; x, y) \neq 0$;
Run violating pair adaptive search on input $(S, \bar{S}; x, y)$ to obtain outputs $\{s\}, \{t\}$.
Contract variables $s$ and $t$ in the oracle and repeat.
```

Proof of Theorem 2. We analyze Algorithm 3. First assume f is k-$\oplus$-partitionable. By Lemma 20, f only contracts variables s, t that are not split by the partition (otherwise $D_f(\{s\}, \{t\}; x', y')$ always vanishes). Therefore f cannot be contracted down to $k-1$ inputs and the tester accepts with probability one.

Now assume f is ϵ-far from partitionable. We will argue that Algorithm 3 outputs "not partitionable" after $O((n-k+1)(\log n + 1/\epsilon))$ queries in expectation by induction on n. Assume $n \geq k$. By Claim 18, Loop 6 takes $O(1/\epsilon)$ iterations to complete in expectation, and each iteration costs four queries to f. By Lemma 20, line 7 takes another $O(\log n)$ queries. After merging s and t the resulting function on $n-1$ inputs can only be farther from partitionable, so by inductive assumption the expected query complexity $Q(n)$ is at most $Q(n-1) + O(\log n + 1/\epsilon)$. This gives the desired bound. By Markov's inequality, Algorithm 3 makes at most twice this number of queries with probability at least half.

The query complexity can be improved slightly to the stated bound $O((n-k)(\log n + 1/\epsilon) + 1/\epsilon)$ by observing that the violating pair search in line 7 can be bypassed when $n = k$ since a proof of non-partitionability has already been discovered in line 6. ◀

5 Testing ⊗-Partitionability

Recall that a function $f : \mathcal{D} = \mathcal{D}_1 \times \cdots \times \mathcal{D}_n \to \mathcal{R}^k$ is *k-⊗-partitionable* if there exists a k-partition $S_1, \dots, S_k$ of $[n]$, and functions $f_1, \dots, f_k$ such that $f(x) = (f_1(x_{S_1}), \dots, f_k(x_{S_k}))$ for all $x \in \mathcal{D}$.

In this section we present a $O((n/\epsilon)\log^2(n/\epsilon))$-query non-adaptive one-sided error test for ⊗-partitionability (see Theorem 3). We begin with an overview of the construction.

First, some notation. For a function $f : \mathcal{D} \to \mathcal{R}^k$ and a subset $T \subseteq [k]$ we write f_T to refer to the function obtained by projecting the output of f onto the coordinates in T. We often write x_i instead of $x_{\{i\}}$ and f_j instead of $f_{\{j\}}$.

For simplicity, suppose that $k = 2$. Then f is 2-⊗-partitionable if and only if every variable has non-zero influence on at most one of the two coordinates of f. So our task boils down to determining whether there is a variable that is influential in both coordinates of the output of f. The key observation that allows us to find such a coordinate with a small number of queries is that if f is ϵ-far from ⊗-partitionable, then

$$\sum_{i \in [n]} \min(\mathrm{Inf}(i; f_1), \mathrm{Inf}(i; f_2)) \geq \epsilon. \tag{4}$$

Therefore, if f is ϵ-far from ⊗-partitionable, there must be a coordinate that has influence at least ϵ/n in both coordinates. This immediately suggests an $O(n^2/\epsilon)$ query test: for each variable use $O(n/\epsilon)$ queries to determine whether it is influential in both coordinates.

We obtain an improvement in the query complexity by exploiting a trade-off between the number of samples, $i \in [n]$, required to find a variable that is influential in both coordinates, and the number of samples required to certify that a variable is indeed influential in both coordinates.

Given subsets $S_1, \dots, S_k$ of $[n]$, let $\Delta_f(S_1, \dots, S_k)$ be the distance from $f = (f_1, \dots, f_k)$ to the closest function $g = (g_1, \dots, g_k)$ in which g_j does not depend on the inputs in S_j, and define the *influence* of $(S_1, \dots, S_k)$ on f as

$$\mathrm{Inf}(S_1, \dots, S_k; f) = \Pr[f_j(x) \neq f_j(y_{S_j} x) \text{ for some } j],$$

where x, y is an independent pair of inputs.

▶ **Proposition 21.** $\mathrm{Inf}(S_1, \dots, S_k; f) \leq \sum_{i=1}^n \mathrm{Inf}(i; f_{J(i)})$, *where* $J(i)$ *is the set of output coordinates* $j \in [k]$ *for which* S_j *contains* i*.*

Proof. Let E be the event "$f_j(x) \neq f_j(y_{S_j}x)$ for some j". Let h^i be the hybrid input in which $h^i_t = y_t$ for $t \leq i$ and x_t for $t > i$. Then $h^0 = x$ and $h^n = y$. If the event "$f_j(x) \neq f_j(y_{S_j}x)$" occurs, then one of the events "$f_j(h^{i-1}_{S_j}x) \neq f_j(h^i_{S_j}x)$" must occur for some i between 1 and n. By the union bound, $\Pr(E) \leq \sum_{i=1}^n \Pr(E_i)$, where E_i is the event "$f_j(h^{i-1}_{S_j}x) \neq f_j(h^i_{S_j}x)$ for some j." The inputs $h^{i-1}_{S_j}x$ and $h^i_{S_j}x$ are identical unless $i \in S_j$, in which case they differ only in the i-th coordinate where they are independent. Therefore

$$\Pr(E_i) = \Pr[f_j(x) \neq f_j(x^i) \text{ for some } j \in J(i)],$$

where x^i is x with its i-th input resampled independently. The right hand side is precisely the influence of i in $f_{J(i)}$. ◀

▷ Claim 22. $\Delta_f(S_1, \dots, S_k) \leq \mathrm{Inf}(S_1, \dots, S_k; f)$.

Proof. By averaging, there must exist an assignment a to y such that

$$\mathrm{Inf}(S_1, \dots, S_k; f) \geq \Pr[f_j(x) \neq f_j(a_{S_j}x) \text{ for some } j].$$

Define $g_j(x) = f_j(a_{S_j}x)$ (on all inputs). Then g_j does not depend on the inputs in S_j, so

$$\Delta_f(S_1, \dots, S_k) \leq \Pr[f_j(x) \neq g_j(x) \text{ for some } j] = \Pr[f_j(x) \neq f_j(a_{S_j}x) \text{ for some } j] \leq \Pr(E).$$

◁

▷ Claim 23. Let $k \geq 2$ and $f(x_i) = (f_1(x_i), \dots, f_k(x_i))$ be a possibly randomized univariate function. Let d an the output coordinate that maximizes $\mathrm{Inf}(i; f_d)$. There exists a partition $(P, \overline{P})$ of the output coordinates such that $\mathrm{Inf}(i; f_P)$ and $\mathrm{Inf}(i; f_{\overline{P}})$ are both at least $\mathrm{Inf}(i; f_{[n]\setminus\{d\}})/3$.

Proof. Let I_j be the event $f_j(x) \neq f_j(y)$ for random independent x and y. Then $\mathrm{Inf}(i; f_T) = \Pr(\cup_{j\in T} I_j)$. Let $\delta_i = \Pr(\cup_{j\neq d} I_j)$. If $\Pr(I_d) \geq \delta_i/3$ then the partition $(\{d\}, [n] \setminus \{d\})$ satisfies the conclusion. Otherwise, $\Pr(I_j) \leq \delta_i/3$ for all j. Then some partition of type $(I_1 \cup \dots \cup I_j, I_{j+1} \cup \dots \cup I_k)$ works: If j is the first set for which $\Pr(I_1 \cup \dots \cup I_j)$ exceeds $\delta_i/3$, then

$$\Pr(I_1 \cup \dots \cup I_j) \leq \Pr(I_1 \cup \dots \cup I_{j-1}) + \Pr(I_j) \leq 2\delta_i/3.$$

Since

$$\Pr(I_1 \cup \dots \cup I_j) + \Pr(I_{j+1} \cup \dots \cup I_k) \geq \Pr(\cup_j I_j) \geq \delta_i,$$

the event $I_{j+1} \cup \dots \cup I_k$ also has probability at least $\delta_i/3$. ◁

▶ **Lemma 24.** *If* f *is* δ*-far from* $\otimes$*-partitionable then there exist partitions* $(P(1), \overline{P(1)})$, …, $(P(n), \overline{P(n)})$ *of* $[k]$ *such that*

$$\sum_{i=1}^n \min\{\mathrm{Inf}(i; f_{P(i)}), \mathrm{Inf}(i; f_{\overline{P(i)}})\} \geq \frac{\delta}{3}.$$

Proof. Let $j^*(i)$ be the maximizer of $\mathrm{Inf}(i; f_j)$ (breaking ties arbitrarily), and S_j be the set of all i such that $j^*(i) \neq j$. Then $J(i) = \{j \colon i \in S_j\} = [n] \setminus \{j^*(i)\}$. By Proposition 21 and Claim 22, $\delta \leq \Delta_f(S_1, \dots, S_k) \leq \sum \mathrm{Inf}(i; f_{[n]\setminus\{j^*(i)\}})$. By Claim 23 applied to f as a function of x_i only (randomized over the other inputs), $\mathrm{Inf}(i; f_{[n]\setminus\{j^*(i)\}})/3 \leq \min\{\mathrm{Inf}(i; f_{P(i)}), \mathrm{Inf}(i; f_{\overline{P(i)}})\}$. ◀

Algorithm 4 Non-adaptive tester for $\otimes$-partitionability.

```
Oracle: $f : \mathcal{D} = \mathcal{D}_1 \times \cdots \times \mathcal{D}_n \to \mathcal{R}^k$
Input : Proximity parameter $\epsilon$
1 foreach $r \in \{0, \dots, \lceil \log(3n/\epsilon) \rceil\}$ do
2     Let $S \subseteq [n]$ be a set of $3 \cdot \lceil \frac{6n \log(3n/\epsilon)}{2^r \epsilon} \rceil$ indices sampled uniformly at random from $[n]$.
3     foreach $i \in S$ do
4         Sample $3 \cdot 2^{r+1}$ independent pairs of inputs from $\mathcal{D}$.
5         if $\exists$ samples $(x, y), (x', y')$, and $j \neq j' \in [k]$ such that $f_j(x) \neq f_j(y_{\{i\}} x)$ and $f_{j'}(x') \neq f_{j'}(y'_{\{i\}} x')$ then
6             Reject.
7 Accept.
```

Proof of Theorem 3. We show that Algorithm 4 satisfies the statement of the theorem. In each iteration of the outer loop, $O(n/\epsilon \log(n/\epsilon))$ queries are made to f. Thus, in total the algorithm makes $O((n/\epsilon) \log^2(n/\epsilon))$ queries.

The test has perfect completeness because the condition on Line 5 is never triggered if f is a direct product.

We now argue soundness. If f is ϵ-far from being a direct product, then by Lemma 24, for every $i \in [n]$ there exist partitions $(P(i), \overline{P(i)})$ such that

$$\sum_{i \in [n]} M_i \geq \epsilon/3, \tag{5}$$

where $M_i = \min\{\mathrm{Inf}(i; f_{P(i)}), \mathrm{Inf}(i; f_{\overline{P(i)}})\}$.

For $r \in \{0, \dots, \lceil \log(3n/\epsilon) \rceil\}$, let A_r denote the set $\{i \mid M_i \in [1/2^r, 1/2^{r+1})\}$. By (5) and an averaging argument, we know that there exists an ℓ such that $|A_\ell| \geq \lceil \frac{2^\ell \epsilon}{6 \log(3n/\epsilon)} \rceil$. For such an ℓ, we show that the probability that the algorithm rejects in the ℓ-th iteration is at least $2/3$.

Consider the ℓ-th iteration of the outer loop. The probability that no index in A_ℓ is picked at Line 2 is at most $(1 - \frac{|A_\ell|}{n})^{3 \cdot \frac{n}{|A_\ell|}} \leq 1/e^3$.

In an iteration of the inner loop corresponding to an index $i \in A_\ell$, the probability that either $f_{P(i)}(x) = f_{P(i)}(y)$ for all sampled pairs (x, y), or $f_{\overline{P(i)}}(x) = f_{\overline{P(i)}}(y)$ for all sampled pairs (x, y) is at most $2 \cdot (1 - 1/2^{\ell+1})^{3 \cdot 2^{\ell+1}} \leq 2/e^3$. This tells us that the probability that the algorithm rejects in the ℓ-th iteration of the outer loop conditioned on $A_\ell \cap S \neq \emptyset$ is at least $(1 - 2/e^3)$. Since $A_\ell \cap S$ is empty with probability at most $1/e^3$, the probability that the algorithm rejects in the ℓ-th iteration is at least $(1 - 3/e^3) \geq 2/3$. ◀

References

1 Rudolf Ahlswede and Peter Gacs. Spreading of sets in product spaces and hypercontraction of the markov operator. *Ann. Probab.*, 4(6):925–939, December 1976. doi:10.1214/aop/1176995937.

2 Eli Ben-Sasson, Madhu Sudan, Salil Vadhan, and Avi Wigderson. Randomness-efficient low degree tests and short PCPs via epsilon-biased sets. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, STOC '03, pages 612–621, New York, NY, USA, 2003. Association for Computing Machinery. doi:10.1145/780542.780631.

3 Eric Blais. Testing juntas nearly optimally. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31 - June 2, 2009*, pages 151–158, 2009. doi:10.1145/1536414.1536437.

4 M. Blum, M. Luby, and R. Rubinfeld. Self-testing/correcting with applications to numerical problems. In *Proceedings of the Twenty-Second Annual ACM Symposium on Theory of Computing*, pages 73–83, New York, NY, USA, 1990. Association for Computing Machinery. doi:10.1145/100216.100225.

5 Andrej Bogdanov and Gautam Prakriya. Direct sum and partitionability testing over general groups. *Electronic Colloquium on Computational Complexity*, 2020. URL: https://eccc.weizmann.ac.il/report/2020/164.

6 Andrej Bogdanov and Baoxiang Wang. Learning and testing variable partitions. In Thomas Vidick, editor, *11th Innovations in Theoretical Computer Science Conference, ITCS 2020, January 12-14, 2020, Seattle, Washington, USA*, volume 151 of *LIPIcs*, pages 37:1–37:22. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020. doi:10.4230/LIPIcs.ITCS.2020.37.

7 Aline Bonami. Étude des coefficients de Fourier des fonctions de $l^p(g)$. *Annales de l'Institut Fourier*, 20(2):335–402, 1970. doi:10.5802/aif.357.

8 Yotam Dikstein and Irit Dinur. Agreement testing theorems on layered set systems. In David Zuckerman, editor, *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019, Baltimore, Maryland, USA, November 9-12, 2019*, pages 1495–1524. IEEE Computer Society, 2019. doi:10.1109/FOCS.2019.00088.

9 Irit Dinur and Konstantin Golubev. Direct sum testing: The general case. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2019, September 20-22, 2019, Massachusetts Institute of Technology, Cambridge, MA, USA*, pages 40:1–40:11, 2019. doi:10.4230/LIPIcs.APPROX-RANDOM.2019.40.

10 Irit Dinur and David Steurer. Direct product testing. In *IEEE 29th Conference on Computational Complexity, CCC 2014, Vancouver, BC, Canada, June 11-13, 2014*, pages 188–196. IEEE Computer Society, 2014. doi:10.1109/CCC.2014.27.

11 Chandra Nair and Yan Nan Wang. Evaluating hypercontractivity parameters using information measures. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 570–574, 2016.

12 Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, USA, 2014.

13 Rocco A. Servedio, Li-Yang Tan, and John Wright. Adaptivity Helps for Testing Juntas. In David Zuckerman, editor, *30th Conference on Computational Complexity (CCC 2015)*, volume 33 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 264–279, Dagstuhl, Germany, 2015. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.CCC.2015.264.

14 Amir Shpilka and Avi Wigderson. Derandomizing homomorphism testing in general groups. *SIAM J. Comput.*, 36(4):1215–1230, 2006. doi:10.1137/S009753970444658X.