# Partial tests, universal tests and decomposability*

Eldar Fischer† Yonatan Goldhirsh‡ Oded Lachish§

July 18, 2014

**Abstract**

For a property $P$ and a sub-property $P'$, we say that $P$ is *$P'$-partially testable with $q$ queries* if there exists an algorithm that distinguishes, with high probability, inputs in $P'$ from inputs $\epsilon$-far from $P$, using $q$ queries. Some natural properties require many queries to test, but can be partitioned into a small number of subsets for which they are partially testable with very few queries, sometimes even a number independent of the input size.

For properties over $\{0,1\}$, the notion of being thus partitionable ties in closely with Merlin-Arthur proofs of Proximity (MAPs) as defined independently in [15]; a partition into $r$ partially-testable properties is the same as a Merlin-Arthur system where the proof consists of the identity of one of the $r$ partially-testable properties, giving a 2-way translation to an $O(\log r)$ size proof.

Our main result is that for some low complexity properties a partition as above cannot exist, and moreover that for each of our properties there does not exist even a single sub-property featuring both a large size and a query-efficient partial test, in particular improving the lower bound set in [15]. For this we use neither the traditional Yao-type arguments nor the more recent communication complexity method, but open up a new approach for proving lower bounds.

First, we use entropy analysis, which allows us to apply our arguments directly to 2-sided tests, thus avoiding the cost of the conversion in [15] from 2-sided to 1-sided tests. Broadly speaking we use "distinguishing instances" of a supposed test to show that a uniformly random choice of a member of the sub-property has "low entropy areas", ultimately leading to it having a low total entropy and hence having a small base set.

Additionally, to have our arguments apply to adaptive tests, we use a mechanism of "re-arranging" the input bits (through a decision tree that adaptively reads the entire input) to expose the low entropy that would otherwise not be apparent.

We also explore the possibility of a connection in the other direction, namely whether the existence of a good partition (or MAP) can lead to a relatively query-efficient standard property test. We provide some preliminary results concerning this question, including a simple lower bound on the possible trade-off.

The positive trade-off result is through the construction of a "universal tester" that works the same for all properties admitting a restricted test. Our tester is very related to the notion of sample-based testing (for a non-constant number of queries) as defined by Goldreich and Ron in [14]. In particular it partially addresses some of the questions raised by [14].

*The research leading to these results has received funding from the European Union's - Seventh Framework Programme [FP7/2007-2013] under grant agreement n 202405 (PROPERTY TESTING).

†Department of Computer Science, Technion, Haifa 32000, Israel. eldar@cs.technion.ac.il

‡Department of Computer Science, Technion, Haifa 32000, Israel. jongold@cs.technion.ac.il

§Birkbeck, University of London, London, UK. oded@dcs.bbk.ac.uk

# 1 Introduction

Property Testing deals with randomized approximation algorithms that operate under low information situations. Formally, we deal with objects from some universe $U$ parametrized by an integer $n$, usually $\Sigma^n$ where $\Sigma$ is some finite alphabet; with a notion of *distance* between two objects in $U$, usually the Hamming distance; and with a notion of a *query* to an object in $U$, usually corresponding to retrieving $x_i$ for an index $i \in \{1, \ldots, n\}$.

**Definition 1.1** (Testable property)**.** *Let $P \subseteq \{0,1\}^n$. We say that $P$ is* testable with $q$ queries *if there exists an algorithm $A$ that gets as input a parameter $\epsilon > 0$ and query access to an input string $x \in \{0,1\}^n$ and outputs* accept *or* reject *such that:*

- *If $x \in P$, then $A$ accepts with probability at least $2/3$.*
- *If $d(x, P) > \epsilon$, then $A$ rejects with probability at least $2/3$.*

*If furthermore all queries performed to the input can be decided before any of them are made, then the algorithm is* non-adaptive*, and otherwise it is* adaptive*. If we require that whenever $x \in P$, then the algorithm accepts with probability* 1*, then the algorithm is* 1-sided*, and otherwise it is* 2-sided*.*

Property Testing was first addressed by Blum, Luby and Rubinfeld [7], and most of its general notions were first formulated by Rubinfeld and Sudan [25]. The first investigated properties were mostly of an algebraic nature, such as the property of a Boolean function being linear. The first investigation of combinatorial properties and the formal definition of testability was by Goldreich, Goldwasser and Ron [13]. Since then Property Testing has attracted significant attention. For surveys see [10, 22, 23].

When proving that testing a property requires many queries, one might ask "how strong is this requirement?", which can be illustrated with an example. Alon et. al. [3] studied the testability of formal languages, and proved that the language $L = \{uu^R vv^R | u, v \in \{0,1\}^*\}$ requires at least $\Omega(\sqrt{n})$ queries to test (formally, the property $L \cap \{0,1\}^n$ requires that many queries to test). Informally, one may say that the "reason" for this language being untestable is the difficulty in guessing the length of $uu^R$. This can be made formal by considering the languages $L_i = \{uu^R vv^R | u, v \in \{0,1\}^*, |u| = i\}$, which form a partition of $L$. A simple sampling algorithm can perform $O(\epsilon^{-1})$ queries to an input and distinguish between inputs in $L_i$ and inputs $\epsilon$-far from $L$. It is also important to note that $|L \cap \{0,1\}^n| = 2^{\Theta(n)}$, but its partition $L_0 \cap \{0,1\}^n, \ldots, L_n \cap \{0,1\}^n$ is only to a number of subsets linear in $n$.

This phenomenon is not unique to the language considered by Alon et. al. Another example is that of graph isomorphism, first considered in the property testing framework by Alon et. al. [2] (and later by Fischer and Matsliah [11]), and shown to require at least $\Omega(n)$ queries to test. In this setting we consider a pair of unknown graphs given by their adjacency matrices, and we are charged with distinguishing the case where they are isomorphic from the case where more than $\epsilon n^2$ of their edges must be changed to make them isomorphic. In this case, the size of the property is $2^{\Theta(n^2)}$, and we can partition the property into $n!$ properties $\{P_\pi | \pi \in S_n\}$, each defined by $P_\pi = \{(G_1, G_2) | \pi(G_1) = G_2\}$, such that a sampling algorithm can perform $O(\epsilon^{-1})$ queries to an input and distinguish between inputs in $P_\pi$ and inputs $\epsilon$-far from the original property.

Thus it is tempting to ask whether this is a general phenomenon. Can any property $P$ be partitioned into $k = |P|^{o(1)}$ properties $P_1, \ldots, P_k$ such that the task of distinguishing inputs in $P_i$ from inputs far from $P$ can be performed with a number of queries that depends only on $\epsilon$?

This question has a strong connection, in fact a near-equivalence, with the notion of a MAP as defined by an independent work of Gur and Rothblum [15]. They define a MAP (Merlin-Arthur proof of Proximity) as a testing algorithm that first read a "proof string" in whole, and uses it to test the given input. The requirement is that an input in $P$ will have some corresponding proof that causes high probability acceptance, while for $\epsilon$-far inputs for every proof there will be a high probability rejection. The connection to our framework is that the proof corresponds to the representation of the alleged $i$ such that the input is in $P_i$, making the required proof length equal to $\lceil \log k \rceil$ for the optimal $k$.

The main result of the present paper is to prove that this is not always the case. In fact, there exist properties for which any such partition must be to a number of subsets exponential in $n$ (and equivalently does not admit a MAP with an $o(n)$ proof size for testing with a number of queries independent of $n$).

To prove this result we in fact show the non-existence of a strictly weaker testing scenario, that would correspond to being able to test just for the biggest $P_i$ in the alleged partition.

**Definition 1.2** (Partially testable property)**.** *Let $P \subseteq \{0,1\}^n$ and $P' \subseteq P$. We say that $P$ is* $P'$-partially testable with $q$ queries *if there exists an algorithm $A$ that gets as input a parameter $\epsilon > 0$ and query access to an input string $x \in \{0,1\}^n$ and outputs* accept *or* reject *such that:*

- *If $x \in P'$, then $A$ accepts with probability at least $2/3$.*
- *If $d(x,P) > \epsilon$, then $A$ rejects with probability at least $2/3$.*

*If furthermore all queries performed to the input can be decided before any of them are made, then the algorithm is* non-adaptive*, and else it is* adaptive*.*

Obviously, if $P$ is testable with $q$ queries, then for any subset $P' \subseteq P$ it is $P'$-partially testable with the same number of queries. On the other hand, for any property $P$ and any element $x \in P$, we have that $P$ is $\{x\}$-partially testable with $O(\epsilon^{-1})$ queries.

The partitions described above are in fact partitions of $P$ into subsets $P_1, \ldots, P_k$ such that $P$ is $P_i$-partially testable for every $1 \leq i \leq k$. If there exists such a partition into not too many sets, then there must be at least one set that is relatively large. Our main result shows that there exists a property $P$ for which all subsets $P' \subseteq P$ such that $P$ is $P'$-partially testable are small. In fact, all linear codes with large dual distance define such properties.

**Theorem 1.3.** *Let $C \subseteq \{0,1\}^n$ be a linear code of size $|C| \leq 2^{\frac{1}{64}n}$ and dual distance $\Gamma$. For every $C' \subseteq C$, if $C$ is $C'$-partially testable with $q$ adaptive queries, then $|C'| \leq |C| 2^{-\Theta(\Gamma/q)}$.*

We will first prove, as a warm-up, a weak version of Theorem 1.3 in Section 4 which will apply for $q$ non-adaptive queries and imply the bound $|C'| \leq |C| 2^{-\Theta(\Gamma/q^3)}$. This proof will use some of the key ideas that will later manifest in the proof of the theorem in its full generality in Section 5.

**Remark 1.4.** *Theorem 1.3 holds for every property $P$ which is $\Gamma$-wise independent. The only use of the linearity of $C$ is in that dual distance $\Gamma$ implies $\Gamma$-wise independence (see Theorem 3.9).*

An important question is the existence of codes with strong parameters. A random linear code $C$ will have $\Gamma = \Theta(n)$ and $|C| = 2^{\Theta(n)}$ with high probability (this is implied by the Gilbert-Varshamov bound [12, 26]; MacWilliams et. al. [19] showed that this can also be obtained by codes which are self-dual and thus also have good distance), and thus by Theorem 1.3 we will have that for any

$C' \subseteq C$ such that $C$ is $C'$-partially testable with $q$ queries, $|C'| \leq |C|2^{-\Theta(n/q)}$. For a constant $q$, this implies that partial testability will only be possible with exponentially small subsets. The best explicit (and low uniform decision complexity) construction known to us is that of [1], which gives $|C| = 2^{\Theta(n)}$ with $\Gamma = \Theta(n/\log n)$, and thus the bound becomes $|C'| \leq n^{O(1)}|C|2^{-\Theta(n/q)}$, which is polynomially worse than the non-explicit bound, but is still a strong upper bound on the size of $C'$.

Theorem 1.3 implies that there exist properties $P$ that require a lot of queries to test, and that every partition of $P$ into subsets $P_1, \ldots, P_k$ such that $P$ is $P_i$-partially testable for every $1 \leq i \leq k$ requires that $k$ will be very big. One might ask if we can prove a converse. That is, if $P$ can be tested with a few queries, can we find such a partition with a small $k$?

**Open Problem 1.5.** *Let $P$ be a property testable with $r$ queries. Is it true that we can partition $P$ into subsets $P_1, \ldots, P_k$ such that $P$ is $P_i$-partially testable with $O(1)$ queries for every $1 \leq i \leq k$ and $k$ is bounded by some moderate function of $r$? What can be said about the converse direction?*

Regarding this problem, that can also be phrased as whether there exists a general trade-off between testing hardness and partitionability to easily partially testable properties. We present some preliminary results that revolve around the much stricter notion of proximity oblivious testing:

**Definition 1.6.** *A* non-adaptive, 1-sided proximity-oblivious $q$-test *for a property $P$ with* detection function $\rho(\epsilon)$ *is an algorithm that makes $q$ non-adaptive queries to the input (i.e. the queries are all made before the answers to them are received), and based on those answers accepts or rejects the input in a way that satisfies the following:*

- *If the input satisfies $P$ then the algorithm accepts with probability* 1.
- *If the input is $\epsilon$-far from $P$, then the algorithm rejects with probability at least $\rho(\epsilon)$.*

*Note that the algorithm is given the input length $n$ in advance, but is not given $\epsilon$. A partial proximity-oblivious $q$-test is defined in the analogous manner.*

The simplest conceivable proximity-oblivious test would be a 2-test, making only 2 queries. Such tests exist for example in some monotonicity testing scenarios. We prove that partitionability into properties that are 2-testable implies a sublinear query test (that is not proximity-oblivious) for the entire property.

**Theorem 1.7.** *Let $P_1, P_2, \ldots, P_k \subseteq \{0,1\}^n$ be properties such that for every $i \in \{1, \ldots, k\}$, $P_i$ has a 1-sided error proximity-oblivious 2-tester with detection function $\rho(\epsilon)$. If $\epsilon > 0$ is such that $\rho(\epsilon/2) > 0$, then for $n$ large enough, as a polynomial function of $1/\rho(\epsilon/2)$, there is a one-sided error non-adaptive $\epsilon$-tester for $P = \bigcup_{i=1}^{k} P_i$ with query complexity $\tilde{O}(n^{2/3}\epsilon^{-1}) \cdot \log(k)$ This is also true if for every $P_i$ we only require a 1-sided error proximity-oblivious $P_i$-partial 2-test for $P$.*

The converse of the above immediately gives an observation interesting enough to state by itself.

**Corollary 1.8.** *If a property $P$ requires $\Omega(n^\beta)$ many queries for some fixed $\beta > 2/3$, then there is no way to partition $P$ into polynomially many properties (even not necessarily disjoint) admitting a 1-sided proximity-oblivious 2-tests (or even the corresponding partial tests).*

Theorem 1.7 is proved using a special test that we call a *universal test*, that works by selecting every index $i$ for querying with probability $\tilde{O}(n^{-1/3}\epsilon^{-1}) \cdot \log(k)$, independently of other indexes. We prove in Theorem 6.8 below that such a kind of test will work for any property admitting a proximity oblivious 2-test, regardless of how that 2-test works. This universal test is very close to what is defined as a *sampling based test* in a new work [14] of Goldreich and Ron. In particular, our proof yields the following corollary, which partially addresses a question from [14] about whether proximity oblivious tests are translatable to sample-based ones:

**Corollary 1.9.** *Let $P$ be a property that has a 1-sided error proximity-oblivious 2-tester with detection function $\rho(\epsilon)$. If $\epsilon > 0$ is such that $\rho(\epsilon/2) > 0$, then for $n$ large enough, as a polynomial function of $1/\rho(\epsilon/2)$, there is a 1-sided error sample based test (see [14], Definition 2.3) with query complexity $\tilde{O}(n^{2/3}\epsilon^{-1}) \cdot \log(k)$.*

For proximity oblivious $q$-tests with $q > 2$ the situation is more complex, and we can only prove an analog of Theorem 6.8 (and by it Theorem 1.7) where the power of $n$ in the query complexity depends (rather badly) on both $q$ and $\rho(\epsilon/2)$.

To formulate the theorem achieving this, we say that a set $R$ of indexes is a *witness against the input* for a property $P$, if the restriction of the input to $R$ is such that it cannot be the restriction of any member of $P$ (or alternatively, this restriction cannot be extended to an alternate input that satisfies $P$).

**Definition 1.10.** *For $\gamma \in (0,1)$, the* $\gamma$-universal sampler *selects a set $R \subseteq [n]$ where, for every $i \in [n]$, $\Pr[i \in R] = n^{-\gamma}$.*

We prove that the above sampling technique, essentially that of a sample-based tester as in [14], is indeed a core of a "universal test" for any property that has a (possibly "unknown") 1-sided proximity-oblivious $q$-test.

**Theorem 1.11.** *For every property $P$ with a proximity oblivious $q$-test with detection function $\rho(\epsilon)$ there exists $\gamma$ depending on $q$ and $\rho(\epsilon/2)$ (for every $\epsilon$), so that for $n$ large enough and every $\epsilon$-far input over $\{0,1\}^n$, the $\gamma$-universal sampler finds a witness against it with probability $1 - o(1)$.*

Its immediate corollary (through standard probability amplification and union bound) gives us a sub-linear query complexity test for any property decomposable into not too many (sub-exponential number of) properties where each of them has a proximity oblivious test, as long as they have the same detection function $\rho(\epsilon)$.

**Corollary 1.12.** *If $P = \bigcup_{i=1}^{\ell} P_i$ is a property such that every $P_i$ has an oblivious 1-sided error (proximity oblivious) $q$-test, all with the same detection function $\rho(\epsilon)$ (but not necessarily the same test), then for $n$ large enough the following is a test for $P$ with $O(\log(\ell)n^{1-\gamma})$ query complexity, where we use the $\gamma$ of Theorem 1.11:*

*Select a set $R \subseteq [n]$ that is the union of $2\log(\ell)$ sets, each chosen according to the $\gamma$-universal sampler. If $|R| > 4\log(\ell)n^{1-\gamma}$ then accept immediately, and otherwise query the input on all indexes of $R$, reject if $R$ is a $P_i$-witness against the input for every $i \in [\ell]$, and accept otherwise.*

Finally, we prove a result in the other direction, hinting that maybe some role for proximity oblivious testing is essential. Using a very simple construction we prove the following:

**Theorem 1.13.** *For every fixed $k$ there is a property $P$, so that $1/5k$-testing $P$ (even adaptively) requires $\Omega(n^{1-1/k})$ queries, while $P$ is still decomposable to at most $n^{k-1}$ many properties so that each of them is even $\epsilon$-testable in itself with $O(1/\epsilon)$ many queries for every $\epsilon$; in fact each of them will have a proximity-oblivious 1-sided $k$-test with the detection function $\rho(\epsilon) = O(k\epsilon)$.*

Until now we discussed the relation of Theorem 1.3 to the impossibility of decomposing a property to testable ones. However, there may be use in its stronger statement of not having even one large sub-property for which there exists an efficient test. The proof of Theorem 1.3 immediately gives the following corollary.

**Corollary 1.14.** *Suppose that $P$ is a property for which $|P| \leq 2^{\frac{1}{64}n}$, and $C$ is any linear code with dual distance $\Gamma$ so that $|P \cap C| \geq |C|2^{-\Gamma/q}$. Then $P$ requires at least $\Theta(q)$ many queries to test (or even $P \cap C$-partially test).*

We conclude this part of the introduction with an open question for which we cannot yet scratch the surface. Theorem 1.3 implies that for some properties, $k$ might be as big as $2^{\Theta(n/q)}$. It is not clear whether this value of $k$ can always be obtained. The trivial upper bound for every property is by partitioning into $2^{n-q}$ subsets of size $2^q$. Are there properties for which this is required?

**Open Problem 1.15.** *Does there exist a property $P$ such that for every $P' \subseteq P$ where $P$ is $P'$-partially testable with $q$ queries we also have $|P'| \leq |P|2^{\Theta(q)-\Theta(n)}$?*

In [15] there is a non-constructive proof (by counting the number of possible algorithms) that there is a property that is not partitionable to less than $2^{\Theta(n)-\Theta(q)}$ properties admitting partial tests with $q$ queries, which would result from a property as above.

## 1.1 Related work

The notion of partial testability, while not defined before, is implicit in previous works on PCPs (Probabilistically Checkable Proofs). The long code tester of Håstad [16] accepts inputs which are codewords in the long code, and rejects inputs which are far from being k-juntas. Since codewords in the long code are junta, this is an instance where the fact that k-juntas are long code-partially testable is used to construct PCPs.

Our notion of a partition is similar to existing notions in computational complexity. For a partition $P = P_1 \cup P_2 \cup \ldots \cup P_k$ where for every $1 \leq i \leq k$, $P$ is $P_i$-partially testable, the designation of $P_i$ can be seen as a "proof" that a certain $x$ is in $P$. If $x \in P$, then there exists some $P_i$ such that $x \in P_i$ and therefore a $P_i$-partial tester for $P$ will accept it with high probability. If $x$ is $\epsilon$-far from $P$, then all $P_i$-partial testers for $P$ will reject it with high probability.

This is similar to the notion of a *Probabilistically Checkable Proof of Proximity (PCPP)*, first introduced by Ben-Sasson et. al. [5]. PCPPs are to property testing as NP is to P. A $q$ query PCPP for a property $P \subset U$ is an algorithm that gets as input $x \in U$ and a *proof of proximity* $\pi \in \{0,1\}^l$. The algorithm must perform at most $q$ queries to $x$ and $\pi$ and fulfill the requirement that if $x \in P$ then there exists a proof $\pi$ that causes the algorithm to accept with high probability, but when $x$ is $\epsilon$-far from $P$ then for any proof $\pi$ the algorithm rejects with high probability. In our setting, the algorithm is allowed free access to a proof of length $l = \log(k)$, but we expect $l$ to be sublinear in the size of $x$. In particular, the property we analyze here cannot have a PCPP with a sublinear length proof. Note that a proof can always be seen as designating a specific subset of the property — the subset of inputs for which this proof is useful.

Rothblum et. al. [24] introduced the notion of an *Interactive Proof of Proximity (IPP)*. In an IPP for a property $P$, the tester can also communicate with a *prover* in addition to querying the input $x$. If $x \in P$ then the prover has a strategy that will cause the tester to accept with high probability. When $x$ is $\epsilon$-far from $P$, the prover cannot make the tester accept with high probability. Rothblum et. al. show that all languages in NC admit such a protocol with $\sqrt{n}$ query and communication complexity and polylog$(n)$ communication rounds. Protocols of this kind are only interesting for the case where the communication complexity is sublinear, or else the prover may just give the input to the tester.

Independently of the present work, Gur and Rothblum [15] weakened the IPP model to create *Merlin-Arthur Proofs of Proximity (MAP)*. Gur and Rothblum define a MAP as a proof-system for a property $P$ where for an input $x$ and a proof $\pi$ the verifier reads the entire proof $\pi$ and queries $q$ bits from the input $x$. If $x \in P$, then there exists a proof $\pi$ such that the verifier accepts with high probability, and if $x$ is far from $P$, then for every proof $\pi$ the verifier rejects with high probability. Since we can trivially set $\pi = x$, the only interesting cases are where the length of $\pi$ is sublinear.

The notion of a MAP with $q$ queries and proofs of length $\ell$ for a property $P$ is equivalent to the existence of $k = 2^\ell$ sets $P_1, \ldots, P_k$ such that $P = P_1 \cup P_2 \cup \ldots \cup P_k$ where for every $1 \leq i \leq k$, $P$ is $P_i$-partially testable with $q$ queries.

Gur and Rothblum give several constructions of properties where a MAP with a sublinear length proof greatly reduces query complexity. Gur and Rothblum also introduce the *Tensor Sum* family of properties, for which they prove that for every constant $\alpha > 0$ there exists an instantiation of Tensor Sum such that any MAP for it that performs $q$ queries must require a proof of length $\Omega\left(\frac{n^{1-\alpha}}{q}\right)$. This bound is slightly weaker than the implication for decomposability of Theorem 1.3 proved in the present paper for our property (however, their property is not a high dual-distance code so our result would not apply directly). There is no known bound on the size of a sub-property of the Tensor Sum properties admitting a partial test, only on decomposability.

Their lower bound is proved by an extension of the communication complexity technique of Brody et. al. [6] to *Merlin-Arthur communication complexity*. First proving a lower bound for 1-sided testing this way, they then use a general conversion technique (at some cost to both proof length and query complexity, see below) to 2-sided testing. Their proof technique is fitting for the MAP setting, but does not apply to partial testing in general. Gur and Rothblum also prove that this trade-off is almost optimal for the Tensor Sum properties.

Additionally, Gur and Rothblum show separations between the power of MAPs and that of IPPs and PCPPs. For their proofs they also show that 2-sidedness may only give a MAP a polylog$(n)$ factor improvement in proof length and query complexity over a 1-sided algorithm. Their result implies a connection also between 1-sided and 2-sided partial testability, though not one that would preserve O(1)-query partial testability.

Regarding the testing versus proof length trade-off question, they show it for the very simple case of "proof-oblivious" testers, i.e. algorithms that make their queries before reading the alleged proof. By contrast, the main difficulty in proving our preliminary trade-off result is exactly that the tests for different $P_i$ could have differing query distributions (even that each of them in itself is proximity oblivious).

Another angle to our methods related to the above trade-off comes from the very recent work of Goldreich and Ron [14]. Their work is centered on what they call *sample-based algorithms*, which are testing algorithms that select all their queries uniformly and independently at random. For a number of queries that is a fixed power of $n$ where $n$ is large enough, this is virtually identical to

the way our universal tests work, where every index is independently chosen to be queried with some fixed probability. Indeed they raised the question whether any property that is testable by a proximity-oblivious $q$-test can also be tested by a sublinear complexity sample-based test, and we give a partial positive answer, for 1-sided error tests ([14] also defines 2-sided error proximity oblivious $q$-tests, which we do not analyze here).

# 2 Plan of the paper

The first major part of the paper is devoted to proving the partial testability lower bound, which is the most "mature" result. After the general preliminaries in Section 3, a warmup proof against non-adaptive testing is found in Section 4, and then additional ideas are incorporated to provide a proof against adaptive testing in Section 5.

The second major part deals with converting proximity-oblivious tests to universal ones that scale well when the property is only decomposable. The conversion from a partition into (proximity-oblivious) 2-testable properties to a test of the whole property is in Section 6, while Section 7 gives the (more expansive) conversion for $q$-testable properties.

Finally, Section 8 contains a property that requires relatively many queries to test for itself, while being partitionable into not too many highly testable properties. For the most part, the sections following the preliminaries can be read individually (Section 7 mildly uses some notions from Section 6).

The rest of this section is devoted to an informal description of the main ideas behind the proofs.

## 2.1 General themes for the partially untestable property

For the proofs of our main result we develop new techniques that are in some ways more flexible than the traditional use of Yao's method for proving property testing lower bounds. We believe that these techniques hold promise for other lower bound situations where using Yao's method seems to hit a wall.

As with Yao's method, we contrast the behavior of a supposed test when it is run over an input chosen according to some distribution over "yes" instances, with its behavior when it is run over an input chosen according to some distribution over "no" instances. However, while in the traditional method these two distributions are chosen based only on the property (and should work against all possible algorithms of a given class), here the distributions are in fact *tailor made* for the specific analyzed algorithm. Note that special care must be taken in the definition of such an input distribution. It may not depend on the "real-time" behavior of the algorithm (i.e. it may not adapt itself to the identity of the random queries that the algorithm has made), and is instead constructed based only on the *description* of the algorithm.

The second theme is the use of *Shannon entropy*. Our goal here is to prove that if $C$ is $C'$-partially testable, then $C'$ cannot be too large. For achieving this we assume that a testing algorithm exists, and then contrast a uniformly random choice of a word in $C'$ with another word chosen from a "dangerous" distribution over words far from $C$. The assumption that the test in fact distinguishes the two distributions allows us to show that a uniformly random choice of a word in $C'$ has low entropy, and hence $C'$ must be small. Using entropy instead of direct counting is crucial to using our main method for obtaining a bound against 2-sided error tests, rather than only 1-sided error ones.

A third theme used in the proof against adaptive algorithms is that of first parsing the input through a specially constructed injective mapping, called a "reader", which is crucial to "exposing" low-entropy portions in this setting. We are in fact considering not just one input distribution, but several of them as the reader is constructed.

### 2.2 Proving a bound against non-adaptive algorithms

The bound against non-adaptive algorithms showcases many of the general themes. A supposed $C'$-partial test with $q$ queries is in essence a distribution over query sets of size $q$, such that with high probability the chosen query set is one that highlights a difference between members of $C'$ and inputs far from being in $C$. As a toy example, assume first that the test is additionally 1-sided, and "well-spread" with respect to the probabilities of querying any particular index. In this case, for every $\epsilon$-far input, the high probability of finding a forbidden substructure (as this is the only way a 1-sided test can reject) translates to having many disjoint $q$-tuples of indexes where in each of them there is a value that a member of $C'$ cannot take (as a hypothetical forbidden structure must exist). This would give a cross product bound on the size of $C'$.

As our tests are not necessarily "well-spread", we will construct a specialized distribution that depends on the specific testing algorithm (but is independent of any particular running instance). For handling 2-sided tests we use a feature of entropy that allows for bounds analogous to combinatorial cross product bounds, namely the subadditivity of the entropy measure.

To construct a "dangerous" distribution over words far from being in $C$, we first take note of the "heavy" indexes, which are those bits of the input that are with high probability part of the query subset of the investigated testing algorithm. There will be only few of those, and our distribution over far words would be that of starting with a restriction of a uniformly random word in $C'$ to the set of heavy indexes, and augmenting it with independently and uniformly chosen values to all other input bits. When contrasted with the uniform distribution over all members of $C'$, we obtain that there must be many query sets that show a distinction between the two distributions over the non-heavy indexes with respect to the heavy ones. This means that the values of the non-heavy indexes in each such query set do not behave like a uniformly independent choice, and thus have a corresponding entropy (conditioned on the heavy index bits) that is significantly less than the maximal possible entropy. Having many such query sets in essence means that we can find many such sets that are disjoint outside the heavy indexes, which in turn leads to an entropy bound by virtue of subadditivity (when coupled with general properties of linear codes).

### 2.3 Proving a bound against adaptive algorithms

An adaptive algorithm cannot be described as a distribution over query sets, but rather as a distribution over small decision trees of height $q$ that determine the queries. Therefore low-entropy index sets cannot be readily found (and in fact do not always exist). To deal with this we employ a new technique, that allows us to "rearrange" the input in a way that preserves entropy, but now admits disjoint low-entropy sets.

This new construction is a *reader*, which in essence is an adaptive algorithm that reads the entire input bit by bit (without repetitions). As this adaptive algorithm always eventually reads the entire input, it defines a bijection between the input to be read and the "reading stream", i.e. the sequence of values in the order that the reader has read them.

The construction of this reader is fully based on the description of the $q$-query adaptive algorithm that $C'$-partially tests for $C$ (again we assume that such an algorithm exists). In fact we contrast the uniform distribution over members of $C'$ with not one but many possible distributions over inputs far from $C$. At every stage we obtain that, as long as our reader has not yet read a large portion of the input, the adaptive test can provide a decision tree over the yet-unread bits that shows a difference between a uniformly random member of $C'$ (conditioned on the values of the bits already read) and an independently uniform random choice of values for the unread bits. Our reader will be the result of "concatenating" such decision trees as long as there are enough unread bits. Thus in the "reading stream" we have sets of $q$ consecutive bits, each with low entropy (as it is distinguishable from independently uniform values). When there are not enough unread bits left, we read all remaining bits arbitrarily, and use general properties of codes to bound the entropy on that final chunk.

The method of constructing a reader not only allows us to do away with the exponential penalty usually associated with moving from non-adaptive to adaptive algorithms, but we additionally obtain better bounds for non-adaptive algorithms as well. This is because a reader can do away also with the penalty of moving from the situation of having many low-entropy query sets to having a family of sets disjoint outside the heavy indexes, in essence by constructing the reader for the uniform distribution over $C'$ based on not one but many "dangerous" input distributions.

### 2.4 Testing decomposable properties through universal testing

Suppose that a property $P$ defined over $\{0,1\}^n$ is decomposable to properties $P_1, \ldots, P_k$, so that each of them is in itself $\epsilon$-testable with $q(\epsilon)$-queries for every $\epsilon > 0$ (the same arguments work also for partial testability, but we restrict the discussion here to proper testability for the sake of explanation). How can we test for all of $P$ at once? The simplest way would be to juxtapose the individual tests for every $P_i$, which would give a test with $O(kq\log(k))$ many queries (accounting also for the necessary probability amplification). However, in our discussion here $k$ rises too fast with $n$, so we would like the dependence on it to be at most polylogarithmic, even at the cost of replacing the "base complexity" $k$ with a value that depends (sublinearly) on $n$.

If the tests for all $P_i$ "behave the same", i.e. have the same query distribution, then instead of querying for every test individually we can do the querying once and feed it to all the tests, and then indeed get a test with $O(q\log(k))$ many queries. This is essentially what is done in the preliminary result from [15]. Our goal here is to replace the original test with a "universal" test that would work for *any* property for which an original test with the specified parameters exist, and then use it instead of the original individual tests.

In our first preliminary result we construct such a test whose number of queries is bounded by a fixed power of $n$, but only if every $P_i$ was testable by the very restricted notion of a 1-sided non-adaptive proximity-oblivious test with 2 queries. Such tests allow for a combinatorial viewpoint through their underlying graphs (where an edge connects two indexes $i, j \in \{1, \ldots, n\}$ if with positive probability the test query set is $\{i, j\}$). This allows for some analysis of the probability of picking a "rejecting edge" when every index ("vertex") is picked and queried with probability $n^{-\beta}$ for an appropriate constant $\beta$. The hard part in the proof is when the test has some "heavy indexes", corresponding to high degree vertices.

Our second result handles proximity-oblivious $q$-tests for any fixed $q$, but unlike the first result, also the power of $n$ in the resulting test depends on $\epsilon$. We essentially make sure that the sampling is "forceful" enough so that any small "erroneous fragment" of the input cannot "propagate" much if

it is altered (the test will detect all possible alterations with large propagations, so such alterations will be forbidden). This in turn allows us to analyze $1/\rho(\epsilon/2)$ many $\epsilon/2$-far inputs derived from the original input, showing that unless the universal test works, they cannot be all rejected by the original test. This is what causes a dependency on $\rho(\epsilon/2)$ of the power of $n$.

### 2.5 A non-testable property that is decomposable to testable ones

In the introduction, the property of being a concatenation of two palindromes was mentioned as one that requires $\Omega(\sqrt{n})$ many queries to test, while being decomposable to $O(n)$ many testable properties (in fact properties admitting a proximity oblivious 2-test). The basic idea from this property is carried over to the properties constructed here. A parity condition ensures that instead of having to correlate two strings (an alleged palindrome and its reverse), we would have to correlate $k$ strings, increasing the bound from $\Omega(\sqrt{n})$ to $\Omega(n^{1-1/k})$. As these $k$ strings are allowed to "slide" relative to each other, the number of properties that we decompose to would be $O(n^{k-1})$, each one corresponding to a fixing of the locations of the strings.

## 3 Preliminaries

Below we introduce the reader to some basic definitions and results regarding entropy and the dual distance of codes. We refer the reader who is interested in a more thorough introduction of entropy to [8, Chapter 2].

First, we introduce a standard notion of distance between distributions.

**Definition 3.1** (Total variation distance)**.** *Let $p$ and $q$ be two distributions over the domain $\mathcal{D}$. The total variation distance between $p$ and $q$ is defined to be $d_{TV}(p,q) = \frac{1}{2}\sum_{i\in\mathcal{D}} |p(i) - q(i)|$.*

We now introduce the notion of the entropy of a random variable, the entropy of a random variable conditioned on another one, and two well-known lemmas.

**Definition 3.2** (Entropy)**.** *Let $X$ be a random variable over the domain $\mathcal{D}$. The* entropy of $X$ *is defined to be $H[X] = -\sum_{i\in\mathcal{D}} \Pr[X=i]\log(\Pr[X=i])$.*

**Definition 3.3** (Conditional entropy)**.** *Let $X$ and $Y$ be random variables over the domain $\mathcal{D}$. The* entropy of $X$ conditioned on $Y$ *is defined to be $H[X|Y] = \sum_{y\in\mathcal{D}} \Pr[Y=i]H[X|Y=y]$.*

**Lemma 3.4** (The chain rule)**.** *Assume that $X$ and $Y$ are random variables. The entropy of the combined state determined by both random variables is denoted $H[X,Y]$. This quantity obeys the chain rule $H[X,Y] = H[X|Y] + H[Y]$.*

**Lemma 3.5** (Subadditivity)**.** *If $X$ and $Y$ are random variables, then $H[X,Y] \leq H[X] + H[Y]$.*

The total variation distance is not a natural fit to the context of entropy. A more fitting notion of distance between distributions is divergence (also known as the Kullback-Liebler divergence [17]).

**Definition 3.6** (Divergence)**.** *Let $p$ and $q$ be two distributions over $\mathcal{D}$. The* divergence of $q$ from $p$ *is defined to be $D(p\|q) = \sum_{i\in\mathcal{D}} p(i)\log\left(\frac{p(i)}{q(i)}\right)$.*

Fortunately, divergence and total variation distance are related via Pinsker's inequality. This was originally proved with worse bounds by Pinsker [20] and seen many subsequent improvements, the current definitive version being that of Reid and Williamson [21].

**Lemma 3.7** (Pinsker's inequality)**.** *Assume that $p$ and $q$ are two distributions over the domain $\mathcal{D}$. The total variation distance between $p$ and $q$ is related to the divergence of $q$ from $p$ by the inequality $\sqrt{\frac{1}{2}D(p||q)} \geq d_{TV}(p,q)$.*

We will actually be using a simpler corollary of it.

**Lemma 3.8** (Corollary of Pinsker's inequality)**.** *Assume that $X$ is a random variable distributed according to $p$ over $\mathcal{D}$, and denote the uniform distribution over $\mathcal{D}$ by $p_u$. The entropy of $X$ is related to its total variation distance from the uniform distribution by $H[X] \leq \log(|\mathcal{D}|) - 2(d_{TV}(p,p_u))^2$.*

*Proof.*

$$H[X] = -\sum_{i\in\mathcal{D}} \Pr[X=i]\log(\Pr[X=i])$$

$$= -\sum_{i\in\mathcal{D}} \Pr[X=i]\log(\Pr[X=i]\cdot\frac{1}{|\mathcal{D}|}\cdot|\mathcal{D}|)$$

$$= -\sum_{i\in\mathcal{D}} \Pr[X=i]\log(\frac{1}{|\mathcal{D}|}) - \sum_{i\in\mathcal{D}} \Pr[X=i]\log(\Pr[X=i]\cdot|\mathcal{D}|)$$

$$= \log(|\mathcal{D}|) - D(p\|p_u) \leq \log(|\mathcal{D}|) - 2(d_{TV}(p,p_u))^2$$

Where the last step follows from Pinsker's inequality. □

Let $x \in \{0,1\}^n$ and $J \subseteq [n]$. We use $x[J]$ to denote the restriction of $x$ to the indices in $J$. That is, the vector $< x_j >_{j\in J}$. When $C \subseteq \{0,1\}^n$ we use $C[J] = \{x[J] | x \in C\}$.

Let $C \subseteq \{0,1\}^n$. We denote by $U(C)$ the uniform distribution over $C$. In accordance with the notation above, when $X \sim U(C)$, $X[J]$ denotes the random variable obtained by drawing uniformly from $C$ and then restricting to the indices in $J$. As a shorthand we use $U(C)[J]$ for the distribution of $X[J]$. We use $U_J(C)$ to denote the result of first drawing a vector $x$ according to $U(C)$, and then replacing $x\,[[n] \setminus J]$ with a uniformly random vector in $\{0,1\}^{n-|J|}$. In particular, in many cases we will take $C$ to be a singleton, in which case we drop the curly braces and denote this probability distribution by $U_J(x)$.

We will make inherent use of the following result, which can be found e.g. in [18, Chapter 1, Theorem 10].

**Lemma 3.9.** *Let $C$ be a linear code with dual distance $\Gamma$. If $J \subseteq [n]$ is such that $|J| < \Gamma$ and $X \sim U(C)$, then $X[J]$ is distributed uniformly over $\{0,1\}^{|J|}$.*

We will also need the fact that a mostly random input is far from a code with high probability.

**Lemma 3.10.** *Let $C \subseteq \{0,1\}^n$ such that $|C| \leq 2^{\frac{1}{64}n}$, $\epsilon < 1/8$, and let $J \subseteq [n]$ be such that $|J| \leq n/2$. $X \sim U_J(C)$ is $\epsilon$-far from $C$ with probability $1 - o(1)$. Furthermore, this is still true when conditioned on any value of $X[J]$.*

*Proof.* By Chernoff bounds, the probability that a random element $X \sim U_J(C)$ will agree with $c \in C$ in more than $(1-\epsilon)n$ coordinates is at most $\exp\left(-n(1/4-\epsilon)^2\right)$. Taking the union bound over all $c \in C$ gives us $|C| \cdot \exp\left(-n(1/4-\epsilon)^2\right) = o(1)$. Since this calculation assumes that $X[J]$ always agrees with $c[J]$, it holds when conditioned on any value of $X[J]$. □

Finally, we will also need to use Lemma 3.9 to help us calculate the entropy of uniform random variables in codes.

**Lemma 3.11.** *Let $C$ be a code with dual distance $\Gamma$, $J \subseteq [n]$ such that $|J| \leq \Gamma$, $C' \subseteq C$ and $X \sim U(C')$. Then $H[X|X[J]] \leq \log|C| - |J|$. Furthermore, this is true when conditioned on any particular value of $X[J]$.*

*Proof.* We can partition $C$ according to the values of the bits in $J$:

$$C = \bigcup_{z \in \{0,1\}^{|J|}} \{c \in C | c[J] = z\}$$

By Lemma 3.9, all sets on the right hand side are of size $2^{-|J|}|C|$. Obviously, for all $z \in \{0,1\}^{|J|}$, we have $\{c' \in C' | c'[J] = z\} \subseteq \{c \in C | c[J] = z\}$, simply because $C' \subseteq C$. Thus for every $x \in C'[J]$, we have that

$$H[X|X[J] = x] \leq \log|\{c' \in C' | c'[J] = z\}| \leq \log|\{c \in C | c[J] = z\}|.$$

This completes the "furthermore" part of the lemma. To obtain the non-conditioned version, note that by the definition of conditional entropy,

$$H[X|X[J]] = \mathrm{E}_{x \sim U(C')[J]} H[X|X[J] = x] \leq \log\left(2^{-|J|}|C|\right) = \log|C| - |J|.$$

□

We note (and use throughout) that trivially $H[X|X[J]] = H[X[\{1,\ldots,n\} \setminus J]|X[J]]$.

# 4 Nonadaptive lower bound

In this section we prove Theorem 1.3 for the case of a non-adaptive tester and with slightly worse quantitative bounds. For the rest of this section, set $C \subset \{0,1\}^n$ to be a code with dual distance $\Gamma$ and $|C| \leq 2^{\frac{1}{64}n}$. Set $\epsilon < 1/8$ and assume that $C$ is $C'$-partially testable for $C' \subseteq C$ with $q$ non-adaptive queries.

Next we define a non-adaptive tester for a property. This definition is consistent with the standard one.

**Definition 4.1** (Non-adaptive property tester)**.** *A non-adaptive $\epsilon$-*tester *for a code $C \subseteq \{0,1\}^n$ with query complexity $q(\epsilon, n)$ is defined by a collection of query sets $\{Q_i\}_{i \in I}$ of size $q$ together with a predicate $\pi_i$ for each query set and a distribution $\mu$ over $I$ which satisfies:*

- *If $x \in C$, then with probability at least $2/3$ an $i \in I$ is picked such that $\pi_i(x[Q_i]) = 1$.*
- *If $d(x, C) > \epsilon$, then with probability at least $2/3$ an $i \in I$ is picked such that $\pi_i(x[Q_i]) = 0$.*

*For a $C'$-partial tester the first item must hold only for $x \in C'$.*

Set a non-adaptive tester for $C'$, and let $\{Q_i\}_{i \in I}$ be its query sets.

We will be interested only in those query sets which are useful for telling a random element in $C'$ from a mostly random element in $\{0,1\}^n$.

**Definition 4.2** ($J$-Discerning query set)**.** *Let $J \subseteq [n]$ be such that $|J| \leq n/2$. A query set $Q_i$ is a* $J$-discerning *set if $d_{TV}(U(C')[Q_i], U_J(C')[Q_i]) \geq 1/8$.*

Next we prove that a tester must have a lot of such good query sets.

**Lemma 4.3.** *Set $J \subseteq [n]$ such that $|J| \leq n/2$. With probability at least $1/9$ the query set $Q_i$ picked by the tester is a $J$-discerning set.*

*Proof.* Assume the contrary, that is, that with probability greater than $8/9$ the query set $Q_i$ picked by the tester is such that $d_{TV}(U(C')[Q_i], U_J(C')[Q_i]) < 1/8$.

Thus for every such $Q_i$,

$$|\Pr_{U(C')[Q_i]}[\text{tester accepts}] - \Pr_{U_J(C')[Q_i]}[\text{tester accepts}]| < 1/8.$$

For the case where the query set picked is not discerning, which occurs with probability smaller than $1/9$, we have no bound (better than 1) on the difference in probability.

Overall, over the randomness of the tester,

$$|\Pr_{U(C')}[\text{tester accepts}] - \Pr_{U_J(C')}[\text{tester accepts}]| < 8/9 \cdot 1/8 + 1/9 = 2/9.$$

But by the correctness of the tester and Lemma 3.10, we arrive at $\Pr_{U(C')}[\text{tester accepts}] \geq 2/3$ and $\Pr_{U_J(C')}[\text{tester accepts}] \leq 1/3$, a contradiction. □

We will later want to construct a collection of $J$-discerning sets disjoint outside of a small fixed portion of the input. Towards this end we prove that $J$-discerning sets show difference between an element in $C'$ and a mostly random element in $\{0,1\}^n$ even when we only look outside of $J$.

**Lemma 4.4.** *Assume that $Q_i$ is a $J$-discerning set, draw $Z \sim U(C')[J]$ and then draw $X \sim U(C')[Q_i]$ conditioned on $X[J] = Z$. With probability at least $1/15$ (taken over the choice of $Z$), the distribution of $X[Q_i \setminus J]$ is $1/16$-far from $U(\{0,1\}^{|Q_i \setminus J|})$.*

*Proof.* First note that the distance between $U(C')[Q_i]$ and $U_J(C')[Q_i]$ is the expectation over $Z$ of the distance of $X[Q_i \setminus J]$ from $U(\{0,1\}^{|Q_i \setminus J|})$, conditioned on $X[J] = Z$. By definition, that is at least $1/8$. By simple probability bounds, with probability at least $1/15$, $Z$ is such that the distance of $X[Q_i \setminus J]$ from $U(\{0,1\}^{|Q_i \setminus J|})$ conditioned on $X[J] = Z$ is at least $1/16$. □

However, total variation distance is not very handy for counting. We now use Lemma 3.8 to transform our total variation bounds into "entropy loss" bounds.

**Lemma 4.5.** *If $Q_i$ is a $J$-discerning set and $X \sim U(C')[Q_i]$, then $H[X[Q_i \setminus J]|X[J]] \leq |Q_i \setminus J| - 0.0005$.*

*Proof.* Let $L \subseteq \{0,1\}^{|J|}$ be the set of values $z \in \{0,1\}^{|J|}$ such that when drawing $X \sim U(C')[Q_i]$ conditioned on $X[J] = z$, the distribution of $X[Q_i \setminus J]$ is $1/16$-far from $U(\{0,1\}^{|Q_i \setminus J|})$.

Since the entropy is non-negative, we can upper bound

$$H[X[Q_i \setminus J]|X[J]] \leq \sum_{z \in L} \Pr_{Z \sim U(C')[J]}[Z = z] H[[Q_i \setminus J]|X[J] = z] + \sum_{z \in \{0,1\}^J \setminus L} \Pr_{Z \sim U(C')[J]}[Z = z]|Q_i \setminus J|.$$

To treat the first summand on the right hand side, we invoke Lemma 3.8 to obtain

$$H[[Q_i \setminus J]|X[J] = z] \leq |Q_i \setminus J| - 0.007.$$

Overall we get

$$\sum_{z \in L} \Pr_{Z \sim U(C')[J]}[Z = z] H[[Q_i \setminus J]|X[J] = z] + \sum_{z \in \{0,1\}^J \setminus L} \Pr_{Z \sim U(C')[J]}[Z = z]|Q_i \setminus J| \leq |Q_i \setminus J| - 0.0005.$$

□

Next, we would try to cover the indices in $[n]$ with as many discerning sets as possible. We will need these sets to be disjoint outside a not-too-big set, so that the "entropy loss" could be aggregated. This set of "bad" indices will be the set of bits read by the tester with the highest probability.

**Definition 4.6.** *Define* $B = \{k \in [n] | \Pr_{Q \sim \mu}[k \in Q] \geq \frac{2q}{\Gamma}\}$.

**Observation 4.7.** $|B| \leq \Gamma/2 \leq n/2$. *Therefore Lemma 3.10 holds with* $J = B$.

Now we can prove that we can find many $B$-discerning sets which are disjoint outside of $B$.

**Lemma 4.8.** *There exists a set* $I_D$ *such that:*

- *For all* $i \in I_D$, $Q_i$ *is a* $B$*-discerning set*
- *For all* $i, j \in I_D$, $Q_i \setminus B$ *and* $Q_j \setminus B$ *are disjoint*

$D = \cup_{i \in I_D}(Q_i \setminus B)$ *satisfies* $\Gamma/2 \geq |D| \geq \frac{\Gamma}{18q^2}$. *Additionally,* $|I_D| \geq \frac{\Gamma}{18q^3}$.

*Proof.* We construct the set $I_D$ greedily. Suppose that we have discerning sets covering $k$ bits that are disjoint outside of $B$. Choose a set randomly using the tester's distribution conditioned on it being $B$-discerning. By Lemma 4.3, this increases the probability of every query set, and every bit to be in a query set, by at most 9. By the definition of $B$, if we choose a query set randomly using the tester's distribution, the probability that it intersects our already covered bits outside of $B$ is at most $9\frac{2q^2}{\Gamma}k$. As long as this number is smaller than 1, such a set exists. Therefore, as long as $k < \frac{\Gamma}{18q^2}$ we have a set to add, leading to the bound. To get the upper bound on $|D|$ we can just stop the process before $D$ gets too big.

The lower bound on the size of $I_D$ follows from the lower bound on the size of $D$. □

Finally, we are ready to calculate the entropy of a uniformly random codeword from $C'$. We use the chain rule to split this into calculating the entropy of the bits in $B$, the entropy of the bits in $D$ conditioned on the bits of $B$, and the entropy of everything else conditioned on the bits in $D \cup B$.

**Lemma 4.9.** *If* $X \sim U(C')$, *then* $H[X] \leq \log |C| - 0.0005\frac{\Gamma}{18q^3}$

*Proof.* First, by the chain rule for entropy and the fact that $D \setminus B = D$,

$$H[X] = H[X|X[D \cup B]] + H[X[D]|X[B]] + H[X[B]]$$

We proceed by bounding each element in the sum. First, trivially:

$$H[X[B]] \leq |B|$$

Next, invoke Lemma 3.11 for $D \cup B$, since $|D \cup B| \leq \Gamma$. This gives us:

$$H[X|X[D \cup B]] \leq \log |C| - |D \cup B|$$

Now, recall that $\cup_{i \in I_D}(Q_i \setminus B) = D$. Since these sets are disjoint outside of $B$, we employ subadditivity to get:

$$H[X[D \setminus B]|X[B]] \leq \sum_{i \in I_D} H[X[Q_i \setminus B]|X[B]]$$

Now, since these are all $B$-discerning sets, by Lemma 4.5 we know that for all $i \in I_D$ we have that $H[X[Q_i \setminus B]|X[B]] \leq |Q_i \setminus B| - 0.0005$. By Lemma 4.8 we know that $|I_D| \geq \frac{\Gamma}{18q^3}$. Summing up we get:

$$\sum_{i \in I_D} H[X[Q_i \setminus B]|X[B]] \leq |D| - 0.0005|I_D|$$

$$\leq |D| - 0.0005\frac{\Gamma}{18q^3}$$

That is,

$$H[X[D]|X[B]] \leq |D| - 0.0005\frac{\Gamma}{18q^3}$$

Summing everything up we get the statement of the lemma. □

From this it follows that:

**Theorem 4.10** (Weak form of the main theorem)**.** *Let $C' \subseteq C$, if $C$ is $C'$-partially testable with $q$ non-adaptive queries, then*

$$|C'| = 2^{H[X]} \leq |C|2^{-0.0005\frac{\Gamma}{18q^3}}.$$

# 5 Adaptive lower bound

In this section we prove Theorem 1.3 in its full generality. We start by introducing the mechanism of a *reader*, which allows us to separate the adaptivity and randomness of the algorithm.

**Definition 5.1** (Reader)**.** *A $k$-*reader $r$ *is a sequence $r_0, r_1, \ldots, r_{k-1}$, where $r_i : \{0,1\}^i \to \{1, \ldots, n\}$ satisfy for all $i < j$ and $y \in \{0,1\}^j$ that $r_i(y[\{1, \ldots, i\}]) \neq r_j(y)$.*

Given an input $x \in \{0,1\}^n$, the reader defines a sequence of its bits. This is the *reading* of $x$.

**Definition 5.2** (Reading)**.** *Given $x \in \{0,1\}^n$ and a $k$-reader $r$, the* reading $R_{r(x)}$ *of $x$ according to $r$ is a sequence $y_1, \ldots, y_k$ defined inductively by $y_{i+1} = x_{r_i(y_1,\ldots,y_i)}$. We define $r_i(x)$ to be $r_i(y_1, \ldots, y_i)$. The set of* unread bits $U_{r(x)}$ *is the subset of $\{1, \ldots, n\}$ that did not appear as values of $r_1, \ldots, r_k$ in the reading.*

We can now define an adaptive tester as a distribution over readers and decision predicates.

**Definition 5.3** (Adaptive tester)**.** *An* adaptive $\epsilon$*-tester for a code* $C \subseteq \{0,1\}^n$ *with query complexity* $q = q(\epsilon, n)$ *is defined by a collection of* $q$*-readers* $\{r^i\}_{i \in I}$ *together with predicates* $\pi_i$ *for each reader, and a distribution* $\mu$ *over* $I$ *which satisfies:*

- *For all* $x \in C$, $\Pr_{i \sim \mu}\left[\pi_i(R_{r^i(x)}) = 1\right] \geq 2/3$.
- *For all* $x \in \{0,1\}^n$ *such that* $d(x, C) > \epsilon$, $\Pr_{i \sim \mu}\left[\pi_i(R_{r^i(x)}) = 0\right] \geq 2/3$.

Part of the usefulness of readers is that if we can construct a reader that reads the entire input, then reading the property $C'$ through it preserves its size.

**Observation 5.4.** *If* $r$ *is an* $n$*-reader, then the function mapping every* $x \in \{0,1\}^n$ *to its reading* $R_{r(x)}$ *is a bijection.*

*Proof.* Suppose that $x' \neq x$, and let $i \in \{1, \ldots, n\}$ be the least index such that $x_{r_i(x)} \neq x'_{r_i(x)}$. Such an $i$ must exist since $r$ reads all bits, and $x' \neq x$. Note that $r_i(x) = r_i(x')$, since it is the first bit read to be different (and thus $y_1, \ldots, y_i = y'_1, \ldots, y'_i$). Thus $x_{r_i(x)} \neq x'_{r_i(x')}$ and therefore $R_{r(x)} \neq R_{r(x')}$. □

In light of the above, we will construct an $n$-reader and bound the size of $C'$ when permuted by its reading. However, while the end product of the construction is an $n$-reader, the intermediate steps might not be $k$-readers for any $k$. Thus we need to introduce a more general notion.

**Definition 5.5** (Generalized reader)**.** *A* generalized reader $r$ *is a sequence* $r_0, r_1, \ldots, r_{n-1}$ *where* $r_i : \{0,1\}^i \to \{1, \ldots, n\} \cup \{\star\}$ *satisfy for all* $i < j$ *and* $y \in \{0,1\}^j$ *one of the following*

- $r_i(y[\{1, \ldots, i\}]) \in \{1, \ldots, n\} \setminus r_j(y)$
- $r_i(y[\{1, \ldots, i\}]) = r_j(y) = \star$

*Given a generalized reader* $r$*, a* terminal sequence *in it is* $y \in \{0,1\}^i$ *such that* $r_i(y_1, \ldots, y_i) = \star$*, while* $r_{i-1}(y_1, \ldots, y_{i-1}) \neq \star$ *or* $i = 0$.

If we fix a certain $x \in \{0,1\}^n$, a generalized reader defines a sequence of non-repeating indices that at some point may degenerate to a constant sequence of $\star$. Note that every $k$-reader naturally defines a generalized reader by setting all undefined functions to map everything to $\star$.

It is useful to think of a (possibly generalized) reader as a decision tree. With a generalized reader, we will often want to continue the branches of the tree with another reader. This operation is called *grafting*. We start with the notion of a 0-*branch* and a 1-*branch*.

**Definition 5.6** (0-branch, 1-branch)**.** *Let* $r$ *be a (possibly generalized) reader. The* 0-branch of $r$ *is the reader* $r'$ *defined by* $r'_i(y_1, \ldots, y_i) = r_{i+1}(0, y_1, \ldots, y_i)$*. Similarly, the* 1-*branch of* $r$ *is the reader* $r''$ *defined by* $r''_i(y_1, \ldots, y_i) = r_{i+1}(1, y_1, \ldots, y_i)$.

We can now define grafting, and will do so recursively. Informally, grafting a reader $t$ onto $r$ at $y$ means that at every $\star$ in the decision tree of $r$ that can be reached after reading $y$, we continue the reading according to $t$. That is, this is the process of appending a decision tree $t$ to another decision tree $r$ given a certain history of reads $y$.

**Definition 5.7** (Grafting)**.** *Let* $r$ *and* $t$ *be generalized readers and* $y \in \{0,1\}^i$ *be a terminal sequence in* $r$*. The* grafting of $t$ onto $r$ on the branch $y$ *is a new reader* $r^{t,y}$ *defined as follows.*

- *If* $t_0 \in \{r_0(y_1,\ldots,y_i),\ldots,r_{i-1}(y_1,\ldots,y_i)\}$*, graft the* $y_{t_0}$*-branch of* $t$ *onto* $r$ *at* $y_1,\ldots,y_i$*.*
- *If* $t_0 \notin \{r_0(y_1,\ldots,y_i),\ldots,r_{i-1}(y_1,\ldots,y_i)\}$*, set* $r_i(y_1,\ldots,y_i)=t_0$*, call the new reader* $r'$*, and graft the* 0*-branch of* $t$ *onto* $r'$ *at* $y_0,\ldots,y_i,0$ *and the* 1*-branch of* $t$ *onto* $t$ *at* $y_0,\ldots,y_i,1$*.*

*Repeat the above recursively, with the base case being the grafting of an identically* $\star$ *reader onto* $r$ *by not changing anything.*

Note that the grafting of a generalized reader onto another results in a generalized reader. Note that it is also possible that $r^{t,y}=r$ when all bits that $t$ may read were already read by $r$ according as $y$.

To introduce the notion of a reader that discerns a random input from an input from $C'$, we will first need to formulate a notion of executing a reader, which is inherently adaptive, on a partly random input.

**Definition 5.8** ($J$-Simulation of a reader)**.** *Let* $r$ *be a* $q$*-reader,* $J\subseteq[n]$ *and* $y\in\{0,1\}^{|J|}$*. The* $J$-simulation of $r$ on $y$ *is the distribution* $S(r,y,J)$ *over* $\{0,1\}^q$ *defined to be* $R_{r(x)}$ *where* $x[J]=y[J]$*, and all bits of* $x$ *outside of* $J$ *are picked independently and uniformly at random from* $\{0,1\}$*.*

We now introduce the notion of a reader that discerns a random input from an input from $C'$.

**Definition 5.9** ($J$-Discerning reader)**.** *Let* $r$ *be a (possibly generalized) reader,* $J\subseteq[n]$ *and* $y\in\{0,1\}^{|J|}$*. Let* $x$ *be a uniform random variable in* $\{c\in C'|c[J]=y\}$*. We say that* $r$ *is a* $J$-discerning reader for $y$ *if* $d_{TV}(R_{r(x)},S(r,y,J))\geq 1/8$*.*

Next, we prove that many readers are indeed discerning.

**Lemma 5.10.** *Set* $J\subseteq[n]$ *such that* $|J|\leq n/2$ *and* $y\in\{0,1\}^{|J|}$*. With probability at least* $1/9$ *the* $q$*-reader* $r$ *picked by the tester is* $J$*-discerning for* $y$*.*

*Proof.* Let $r$ be a reader that is not $J$-discerning for $y$. Let $B\sim U_J(y)$ and $G\sim U(\{c\in C'|c[J]=y\})$. Denote by $\pi_r$ the predicate associated with $r$. By our assumption,

$$|\Pr[\pi_r(R_{r(B)})=1]-\Pr[\pi_r(R_{r(G)})=1]|<1/8.$$

Now assume that with probability greater than $8/9$, the $q$-reader picked is not $J$-discerning for $y$. Now consider the difference in acceptance probability when drawing a reader according to $\mu$.

$$|\Pr_{r\sim\mu}[\pi_r(R_{r(B)})=1]-\Pr_{r\sim\mu}[\pi_r(R_{r(G)})=1]|<8/9\cdot 1/8+1/9=2/9.$$

But by Lemma 3.10 and the correctness of the tester, $\Pr_{r\sim\mu}[\pi_r(R_{r(B)})=1]\leq 1/3$, and by the correctness of the tester $\Pr_{r\sim\mu}[\pi_r(R_{r(G)})=1]\geq 2/3$, a contradiction. □

A common operation will be to graft a discerning reader with additional arbitrary bits. This does not cause a discerning reader to stop being one.

**Definition 5.11.** *Let* $r$ *and* $s$ *be generalized readers. We say that* $r$ contains $s$ *if for every* $x\in\{0,1\}^n$*, the sequence of non-*$\star$ *elements in* $R_{s(x)}$ *is a prefix of* $R_{r(x)}$*.*

Note that in particular, whenever we graft $s$ onto $r$ along some branch, we obtain a reader which contains $r$.

**Lemma 5.12.** *Let $r$ and $s$ be generalized readers such that $r$ contains $s$. Let $J \subseteq [n]$ and $y \in \{0,1\}^{|J|}$. If $s$ is a $J$-discerning reader for $y$, then so is $r$.*

*Proof.* Let $B \sim U_J(y)$ and $G \sim U(\{c \in C' | c[J] = y\})$. Consider $R_{r(B)}$. Its outcomes can be partitioned according to their $R_{s(B)}$ prefixes. Thus the probability of every event defined by values of $R_{r(B)}$ can be written as a weighted sum of the probabilities of events defined by values of $R_{s(B)}$. The same is true for $R_{r(G)}$ and $R_{s(G)}$. Therefore $d_{TV}(R_{r(x)}, S(r,y,J)) \geq d_{TV}(R_{s(x)}, S(s,y,J))$. □

To prove that a uniform choice in $C'$ does not have high entropy we graft discerning readers one onto the other. We will want to make sure that all the branches of the decision tree are of the same height throughout the grafting, and thus we define the notion of a *padded grafting*.

**Definition 5.13** ($q$-Padded grafting)**.** *Let $r$ be a generalized reader, $t$ be a $q$-reader and $y \in \{0,1\}^i$ be a terminal sequence in $r$. The* $q$-padded grafting of $t$ onto $r$ on the branch $y$ *is defined by the following process. First, let $r'$ be the grafting of $t$ onto $r$ at the branch $y$. Now perform the following repeatedly: Let $z_1, \ldots, z_j$ with $j < q$ be such that $r'_{i+j-1}(y_1, \ldots, y_i, z_1, \ldots, z_{j-1}) \neq \star$ but $r'_{i+j}(y_1, \ldots, y_i, z_1, \ldots, z_j) = \star$ (or $j = 0$ and $r'_i(y_1, \ldots, y_i) = \star$). Let $k$ be an arbitrary index not in $\{r'_0, \ldots, r'_{i+j-1}(y_1, \ldots, y_i, z_1, \ldots, z_{j-1})\}$, and redefine $r'_{i+j}(y_1, \ldots, y_i, z_1, \ldots, z_j) = k$. Repeat this process as long as such $z_1, \ldots, z_j$ with $j < q$ exist.*

The above is basically grafting additional arbitrary reads, so that the end-result will always read exactly $q$ bits after reading the sequence $y_1, \ldots, y_i$. The next observation together with Lemma 5.12 implies that $q$-padded grafting of a $J$-discerning reader is equivalent to a grafting of some other $J$-discerning reader.

**Observation 5.14.** *Let $r$ be a generalized reader, $t$ a $q$-reader and $y \in \{0,1\}^i$ a terminal sequence in $r$. There exists a reader $s$ containing $t$ such that the $q$-padded grafting of $t$ onto $r$ at $y$ is equivalent to the grafting of $s$ onto $r$ at $y$.*

Now we can finally prove the main lemma, by performing repeated $q$-padded grafting of discerning readers one onto another.

**Lemma 5.15.** *If $X \sim U(C')$, where $C$ is $C'$-partially testable with $q$ queries, then $H[X] \leq \log|C| - \lfloor \frac{1}{32}\Gamma/q \rfloor$*

*Proof.* Let us construct an $n$-reader and consider the entropy of $C'$ when permuted by this reader.

Start with a 0-reader $r^0$. Let $s$ be a $\emptyset$-discerning $q$-reader for the empty word, which must exist since the adaptive tester must pick one with positive probability. Set $r^1$ to be the grafting of $s$ onto $r^0$ on the branch of the empty word.

Assume that we have constructed the $jq$-reader $r^j$. If $jq \geq \Gamma$, graft a reader that reads all remaining bits arbitrarily onto $r^j$ on all branches. Else, perform the following for all branches $y \in \{0,1\}^{jq}$ to obtain $r^{j+1}$ (noting that they are all terminal sequences in $r^j$):

- If there is no member of $C'$ with the reading $R_{r^j(y)}$, perform a $q$-padded grafting of an arbitrary $q$-reader onto $r^j$ at the branch $y$,
- If such a member exists, let $s$ be a $\{r^j_1(y), r^j_2(y), \ldots, r^j_{jq}(y)\}$-discerning reader for $y$. Perform a $q$-padded grafting of $s$ onto $r^j$ at the branch $y$.

Now let $r$ be the resulting $n$-reader, let $r_{R(C')}$ be the image of $C'$ under the reading of $r$, and let $X \sim U(r_{R(C')})$. By Observation 5.4, the distribution of $X$ is the same as starting with a uniformly random member of $C'$ and then taking its reading according to $r$. By the chain rule $H[X] = H[X[\{1,\ldots,\Gamma\}]] + H[X|X[\{1,\ldots,\Gamma\}]]$.

Note that in the case of a word from $C'$, the maximal $j$ in the construction is equal to $\Gamma/q$. By the chain rule we may write

$$H[X[\{1,\ldots,\Gamma\}]] = \sum_{i=1}^{\Gamma/q} H[X[\{(i-1)q+1,\ldots,iq-1\}]|X[\{1,\ldots,(i-1)q-1\}]]$$

and since each sequence of $q$ bits is from the grafting of a reader which is discerning with respect to all the previous ones, we may apply Lemma 3.8 to obtain

$$H[X[\{1,\ldots,\Gamma\}]] = \sum_{i=1}^{\Gamma/q} H[X[\{(i-1)q+1,\ldots,iq-1\}]|X[\{1,\ldots,(i-1)q-1\}]]$$

$$\leq \sum_{i=1}^{\Gamma/q} \left(q - \frac{1}{32}\right) \leq \Gamma - \Gamma/q \cdot \frac{1}{32}$$

.

By Lemma 3.11, $H[X|X[\{1,\ldots,\Gamma\}]] \leq \log|C| - \Gamma$, so by summing it all up we get $H[X] \leq \log|C| - \Gamma/q \cdot \frac{1}{32}$.

□

This gives us Theorem 1.3 in its full generality, as it implies that $|C'| = 2^{H[X]} \leq 2^{-\Gamma/32q} \cdot |C|$.

# 6 Properties with a proximity oblivious 2-testable decomposition

For simplicity of presentation all the proofs here are for a property $P$ which is decomposable to properties $P_1,\ldots,P_\ell$ that in themselves admit a proximity oblivious 2-test rather than just a $P_i$-partial test for $P$. A sketch on how to extend this to the more general setting is found at the end of this section.

**Definition 6.1** ($P$-witness)**.** *Let $P \subseteq \{0,1\}^n$ be a property and $w \in \{0,1\}^n$. A* $P$-witness *against $w$ is a set $Q \subseteq [n]$ such that for every $w' \in \{0,1\}^n$, if $w'_i = w_i$ for every $i \in Q$, then $w' \notin P$.*

The family of witness sets for a specific $w$ is closed to supersets. Note that any 1-sided $q$-test essentially rejects only if their query set is a witness. A proximity oblivious 1-sided test is a non-adaptive one which is also independent of the proximity parameter $\epsilon$, essentially just a probability distributions over query sets of a fixed size $q$. This means that the following definition of a proximity-oblivious test is in fact equivalent to Definition 1.6 from the introduction.

**Definition 6.2** (proximity oblivious test by witnesses)**.** *A* proximity oblivious 1*-sided $q$-test with the* detection function $\rho(\epsilon)$ *is a probability distributions over query sets of a fixed size $q$, so that for every $\epsilon$-far input $w$ (for every $\epsilon$) the probability of obtaining a witness against $w$ is at least $\rho(\epsilon)$.*

**Definition 6.3** (universal sampler)**.** *For parameters $\epsilon, \eta \in (0,1)$, the* $(\epsilon,\eta)$-universal sampler *selects a set $R \subseteq [n]$ where, for every $i \in [n]$, $\Pr[i \in R] = \alpha^3 n^{-1/3}$, with the parameter $\alpha = 8\epsilon^{-1} \log \epsilon^{-1} \cdot \log \eta^{-1} \cdot \log n$.*

*Let $P_1, P_2, \ldots, P_\ell \subseteq \{0,1\}^n$ be properties, each having an oblivious one-sided error 2-tester with the same detection function $\rho(\epsilon)$. Given oracle access to input $w \in \{0,1\}^n$, the* $\epsilon$-universal algorithm *for $\bigcup_{i=1}^{\ell} P_i$, selects a set $R \subseteq [n]$ according to the $(\epsilon, 1/4\ell)$-universal sampler. If $|R| > 2\alpha^3 n^{2/3}$, then it accepts immediately, and otherwise it queries the input on all indices of $R$, rejects if $R$ is a $P_i$-witness against $w$ for every $i \in [\ell]$, and accepts otherwise.*

**Lemma 6.4** (implying Theorem 1.7)**.** *Let $P_1, P_2, \ldots, P_\ell \subseteq \{0,1\}^n$ be properties such that for every $i \in [\ell]$, $P_i$ has a one-sided error oblivious 2-tester with detection function $\rho(\epsilon)$. If $\epsilon > 0$ is such that $\rho(\epsilon/2) > 0$, then for $n$ large enough (as a polynomial function of $1/\rho(\epsilon/2)$) the $\epsilon$-universal algorithm for $\bigcup_{i=1}^{\ell} P_i$ is a one-sided error non-adaptive $\epsilon$-tester for $\bigcup_{i=1}^{\ell} P_i$ with query complexity bound $O(n^{2/3}(\epsilon^{-1} \log \epsilon^{-1} \cdot \log \eta^{-1} \cdot \log n)^3)$.*

To arrive at the theorem, we first need to "thin out" the possible test queries.

**Definition 6.5** ($\epsilon$-trap)**.** *A set $\mathcal{Q}$ of size-2 subsets of $[n]$ is called an* $\epsilon$-trap *for a property $P$, if for every word $w \in \{0,1\}^n$ that is $\epsilon$-far from $P$, there is some set $Q \in \mathcal{Q}$ which is a $P$-witness against $w$.*

**Lemma 6.6.** *If $P$ has a one-sided error oblivious 2-tester with the detection function $\rho(\epsilon)$, then for every $\epsilon$ it has an $\epsilon$-trap $\mathcal{Q}$ with $|\mathcal{Q}| \leq 9n/\rho(\epsilon)$.*

*Proof.* This is immediate from running the 2-tester $9n/\rho(\epsilon)$ many times (so with positive probability it will happen that every possible $\epsilon$-far word is rejected by some iteration of it), and then setting $\mathcal{Q}$ to be the set of all query sets drawn in these iterations. □

We will also use as usual the following triviality.

**Observation 6.7.** *For $n$ larger than some universal constant, the $\epsilon$-universal test will execute the "immediate accept" step (due to $R$ being too large) with probability less than $1/12$.*

Lemma 6.4 and hence Theorem 1.7 now follows by first obtaining $\mathcal{Q}_1, \ldots, \mathcal{Q}_\ell$ as $\epsilon/2$-traps for $P_1, \ldots, P_\ell$ respectively, and then using the union bound for the respective applications of the following statement, which is in some ways the "true theorem" of this section.

**Theorem 6.8.** *Let $\epsilon > 0$, $\eta > 0$, $\mathcal{Q}$ be an $\epsilon/2$-trap for a property $P$, and $w$ be $\epsilon$-far from $P$. For $n$ larger than some polynomial function of $|\mathcal{Q}|/n$, the set $R$ produced by the $(\epsilon,\eta)$-universal sampler is a $P$-witness against $w$ with probability exceeding $1-\eta$.*

**Observation 6.9.**

1. $(1-\alpha^3 n^{-1/3})^{\epsilon n/4} < \eta 2^{-2\epsilon n^{2/3}}/3$,
2. $(1-\alpha^3 n^{-1/3})^{\alpha^{-2} n^{1/3}} < \eta/3n$,
3. $e^{-\epsilon\alpha(\alpha-4)} < \eta$.

From here on we fix $P$ to be a property, $\mathcal{Q}$ to be its $\epsilon/2$-trap, and $w \in \{0,1\}^n$ to be $\epsilon$-far from $P$.

**Definition 6.10** (degree)**.** *For every $i \in [n]$ and $\mathcal{Q}' \subseteq \mathcal{Q}$, we define* $\deg_{\mathcal{Q}'}(i) = |\{Q \in \mathcal{Q}' \mid i \in Q\}|$.

**Definition 6.11** ($\mathcal{W}_w$, $\mathcal{L}_w$ and $\mathcal{M}_w$)**.** *For every $w \in \{0,1\}^n$*

1. *$\mathcal{W}_w$ is the set of all members of $\mathcal{Q}$ that are witnesses against $w$, and for every $i \in [n]$, $\mathcal{W}_w^i$ is the set of all members of $\mathcal{W}_w$ that contain $i$.*
2. $\mathcal{L}_w = \left\{ Q \in \mathcal{W}_w \;\middle|\; \exists j \in Q \text{ s.t. } \deg_{\mathcal{W}_w}(j) > \alpha^{-2} n^{1/3} \right\}$
3. $\mathcal{M}_w = \mathcal{W}_w \setminus \mathcal{L}_w$.

**Definition 6.12** (the $\Rightarrow$ notation)**.** *Let $(i,a),(j,b) \in [n] \times \{0,1\}$ be distinct. We write $(i,a) \Rightarrow (j,b)$ if $\mathcal{Q}$ has no witness against some $w' \in \{0,1\}^n$ such that $w'_j = \neg b$ while $\mathcal{Q}$ has a witness against every $w^* \in \{0,1\}^n$ such that $w^*_i = a$ and $w^*_j = \neg b$.*

**Definition 6.13** (viable sub-string)**.** *Let $B \subset [n]$ be a set of indexes and $\sigma_B : B \to \{0,1\}$. $\sigma_B$ is a* viable sub-string *if there exist no $h \in [n]$, $a \in \{0,1\}$ and $i,j \in B$, that are not necessarily distinct, such that $(i,\sigma_B(i)) \Rightarrow (h,a)$ and $(j,\sigma_B(j)) \Rightarrow (h,\neg a)$, or $(i,\sigma_B(i)) \Rightarrow (h,a)$ and $\mathcal{Q}$ has a witness against every $w^* \in \{0,1\}^n$ such that $w^*_h = a$.*

**Definition 6.14** (witness against sub-string)**.** *Let $B \subset [n]$ and $\sigma_B : B \to \{0,1\}$ be a viable sub-string. $i \in [n]$ is a* witness against $\sigma_B$ *in $w \in \{0,1\}^n$, if $i \in B$ and $w_i \neq \sigma_B(i)$, or if there exists $j \in B$ such that $(j,\sigma_B(j)) \Rightarrow (i,\neg w_i)$.*

**Definition 6.15** ($\mathrm{Inf}_{\sigma_B}$, $\sigma_B^{\mathrm{Inf}}$)**.** *Let $B \subset [n]$ and $\sigma_B : B \to \{0,1\}$ be a viable sub-string. We define $\mathrm{Inf}_{\sigma_B}$ to be the set containing all the possible witnesses against $\sigma_B$. We define $\sigma_B^{\mathrm{Inf}} : \mathrm{Inf}_{\sigma_B} \to \{0,1\}$ so that for every $i \in B$ and $j \in \mathrm{Inf}_{\sigma_B}$, if $(i,\sigma_B(i)) \Rightarrow (j,a)$, then $a = \sigma_B^{\mathrm{Inf}}(j)$.*

**Lemma 6.16.** *Let $B \subset [n]$, $\sigma_B$ be a viable sub-string. For every $w^* \in \{0,1\}$ such that $w^*_i = \sigma_B^{\mathrm{Inf}}(i)$, for every $i \in \mathrm{Inf}_{\sigma_B}$, all of the members of $\mathcal{W}_{w^*}$ are disjoint from $\mathrm{Inf}_{\sigma_B}$.*

*Proof.* Assume for the sake of contradiction that the lemma does not hold. If there exists a member of $\mathcal{W}_{w^*}$ that is contained in $\mathrm{Inf}_{\sigma_B}$, then $\sigma_B$ is not a viable sub-string and hence the contradiction. If $\mathcal{W}_{w^*}$ has a member $\{i,j\}$ such that $i \in \mathrm{Inf}_{\sigma_B}$ and $j \notin \mathrm{Inf}_{\sigma_B}$, then there exists $h \in B$ such that $(h,\sigma_B(h)) \Rightarrow (j,\neg w^*_j)$ (we take the $h$ for which $(h,\sigma_B(h)) \Rightarrow (i,w^*_i)$). This is a contradiction to the definition of $\mathrm{Inf}_{\sigma_B}$ as containing all such $j$. □

**Lemma 6.17.** *Let $w$ be $3\epsilon/4$-far from $P$. If $\mathcal{W}_w \subseteq 2^B$, then $\mathrm{Inf}_{\sigma_B}$ contains at least $\epsilon n/4$ witnesses against $\sigma_B$ for any viable sub-string $\sigma_B : B \to \{0,1\}$.*

*Proof.* Assume for the sake of contradiction that $\mathrm{Inf}_{\sigma_B}$ contains less than $\epsilon n/4$ witnesses against $\sigma_B$. Let $w^* \in \{0,1\}$ be such that $w^*_i = \sigma_B^{\mathrm{Inf}}(i)$ if $i \in \mathrm{Inf}_{\sigma_B}$ and otherwise $w^*_i = w_i$. Obviously, $w^*$ is $3\epsilon/4$-far from $P$.

By Lemma 6.16, $\mathcal{W}_{w^*}$ does not have any sets that intersect $\mathrm{Inf}_{\sigma_B}$. Since $\mathcal{W}_w \subseteq 2^B$, $\mathcal{W}_{w^*} \cap 2^{[n]\setminus \mathrm{Inf}_{\sigma_B}} = \emptyset$. Thus, $\mathcal{W}_{w^*} = \emptyset$ and hence $\mathcal{Q}$ has no witness against $w$. This is a contradiction to $\mathcal{Q}$ being an $3\epsilon/4$-trap for $P$. □

**Lemma 6.18.** *If $|\bigcup_{Q \in \mathcal{W}_w} Q| \leq 2\epsilon n^{2/3}$, then we have that $\Pr[R \text{ is a witness against } w] > 1 - \eta/3$, even if $w$ is only $3\epsilon/4$-far from $P$.*

*Proof.* Let $W = \bigcup_{Q \in \mathcal{W}_w} Q$. By assumption, $|W| \leq 2\epsilon n^{2/3}$. Let $\sigma_W$ be a viable sub-string. Since $\mathcal{W}_w \subseteq 2^W$, by Lemma 6.17, there are at least $\epsilon n/4$ witnesses against $\sigma_W$.

The probability that such a witness is not selected is at most $(1 - \alpha^3 n^{-1/3})^{\epsilon n/4} < \eta 2^{-2\epsilon n^{2/3}}/3$, where the inequality is by Observation 6.9. The lemma follows by the union bound over all viable sub-strings for $W$. □

**Lemma 6.19.** *If $|\bigcup_{Q \in \mathcal{W}_w} Q| > 2\epsilon n^{2/3}$ and $|\bigcup_{Q \in \mathcal{M}_w} Q| < \epsilon n^{2/3}$, then the probability that $R$ is a witness against $w$ is at least $1 - \eta$, for $n$ larger than some polynomial in $|\mathcal{Q}|/n$.*

*Proof.* Observe that $|\mathcal{L}_w| \geq \epsilon n^{1/3}$, because by definition we have $\bigcup_{Q \in \mathcal{W}_w} Q = \bigcup_{Q \in \mathcal{M}_w \cup \mathcal{L}_w} Q$. Let $\mathrm{Piv} \subseteq [n]$ be the set of all $i$ such that $\deg_{\mathcal{W}_w}(i) \geq \alpha^{-2} n^{1/3}$, and $\sigma_{\mathrm{Piv}}$ be such that $\sigma_{\mathrm{Piv}}(i) = \neg w_i$ for every $i \in \mathrm{Piv}$. Note that, for every $i \in \mathrm{Piv}$,

$$\Pr[R \text{ does not contain } j_i \text{ such that } \{i, j_i\} \in \mathcal{W}_w]$$

$$\leq (1 - \alpha^3 n^{-1/3})^{\alpha^{-2} n^{1/3}} < \eta/3n,$$

where the last inequality is by Observation 6.9. Consequently, by the union bound

$$\Pr[\text{for every } i \in \mathrm{Piv},\ \exists j_i \in R \text{ s.t. } \{i, j_i\} \in \mathcal{W}_w] > 1 - \eta/3.$$

When the event above indeed occurs, it is only for $\sigma = \sigma_{\mathrm{Piv}}$ (out of any $\sigma : \mathrm{Piv} \to \{0,1\}$) that it may be the case that $\{j_i : i \in \mathrm{Piv}\}$ is not a witness against $\sigma$. In other words, with probability at least $1 - \eta/3$ we obtain the event that $R$ contains witnesses against all possible assignments to Piv, apart from possibly $\sigma_{\mathrm{Piv}}$. To conclude we partition to two cases that depend on the relationship of $\sigma_{\mathrm{Piv}}$ and $w$.

If $w$ has at least $\epsilon n^{1/3}$ witnesses against $\sigma_{\mathrm{Piv}}$, then the probability that such a witness is not selected is less than $(1 - \alpha^3 n^{-1/3})^{\epsilon n^{1/3}} < \eta/3$, where the inequality is by Observation 6.9. Thus, by the union bound, with probability exceeding $1 - \eta$, $R$ is a witness against $w$ (as it contains witnesses against any possible assignment to Piv).

Assume now that there are less than $\epsilon n^{1/3}$ witnesses against $\sigma_{\mathrm{Piv}}$. Let $w^* \in \{0,1\}^n$ be such that if $i \in \mathrm{Inf}_{\mathrm{Piv}}$, then $w^*_i = \sigma^{\mathrm{Inf}}_{\mathrm{Piv}}(i)$, and otherwise $w^*_i = w_i$. By the triangle inequality, $w^*$ is $3\epsilon/4$-far from $P$ (for $n$ large enough so that $\epsilon n^{1/3} < \epsilon n/4$). By Lemma 6.16, none of the sets in $\mathcal{W}_{w^*}$ intersect $\mathrm{Inf}_{\mathrm{Piv}}$ and hence $\mathcal{W}_{w^*} \subseteq \mathcal{M}_w$ Consequently,

$$|\bigcup_{Q \in \mathcal{W}_{w^*}} Q| < |\bigcup_{Q \in \mathcal{M}_w} Q| \leq 2\epsilon n^{2/3}.$$

Thus, by Lemma 6.18, with probability exceeding $1 - \eta$, $R$ is a witness against $w^*$ and so against $w$ (this case does not even require us to analyze witnesses against the possible assignments to Piv themselves). □

**Lemma 6.20.** *If $|\bigcup_{Q \in \mathcal{M}_w} Q| \geq \epsilon n^{2/3}$, then $\Pr[\mathcal{W}_w \cap 2^R \neq \emptyset] > 1 - \eta$.*

*Proof.* Let $R'$ be a random subset of $R$, where every member of $R$ is in $R'$ independently with probability $\alpha^{-2}$. We observe that, by the definition of $R$, for every $i \in [n]$ independently, we have that $i \in R'$, with probability $\alpha n^{-1/3}$. We next prove that $\Pr[\mathcal{M}_w \cap 2^{R'} \neq \emptyset] > 1 - \eta$, since $R' \subseteq R$ and $\mathcal{M}_w \subseteq \mathcal{W}_w$, this implies the Lemma.

For every integer $i$, let $d(i) = |\{j \in [n] \mid 2^i \leq \deg_{\mathcal{M}_w}(j) < 2^{i+1}\}|$. Let $\Delta$ be the expected number of pairs of distinct $Q, Q' \in \mathcal{M}_w \cap 2^{R'}$ such that $Q \cap Q' \neq \emptyset$. We observe that,

$$\Delta \leq \alpha^3 n^{-1} \sum_{i=1}^{\frac{\log n}{3} - 2\log\alpha} \binom{2^{i+1}}{2} d(i).$$

We observe that, $d(i) \leq |\mathcal{M}_w| 2^{-i+1}$. Plugging this into the above,

$$\Delta < \alpha^3 n^{-1} \sum_{i=1}^{\frac{\log n}{3} - 2\log\alpha} 2^{1+2i} |\mathcal{M}_w| 2^{-i+1} \leq 4\alpha^3 n^{-1} |\mathcal{M}_w| \sum_{i=1}^{\frac{\log n}{3} - 2\log\alpha} 2^i \leq 8\alpha n^{-\frac{2}{3}} |\mathcal{M}_w|. \tag{1}$$

For every $Q \in \mathcal{M}_w$, let $X_Q$ be a random variable that is 1, if $Q \subseteq R'$ and otherwise 0. Let $\mu$ be the expected value of $\sum_{Q \in \mathcal{M}_w} X_Q$. Then,

$$\mu = \frac{\alpha^2 |\mathcal{M}_w|}{n^{\frac{2}{3}}}. \tag{2}$$

Consequently, by (1), (2) and Janson's inequality [4, Part 8],

$$Pr[\mathcal{M}_w \cap 2^{R'} = \emptyset] \leq e^{-n^{-\frac{2}{3}} |\mathcal{M}_w| \alpha(\alpha-4)} < e^{-\epsilon\alpha(\alpha-4)} < \eta,$$

where the second to last inequality follows from $|\mathcal{M}_w| \geq \epsilon n^{\frac{2}{3}}$ and the last from Observation 6.9. □

*Proof of Theorem 6.8.* An $\epsilon$-far $w$ must clearly fall under at least one of Lemma 6.18, Lemma 6.19 and Lemma 6.20. □

We conclude this section with a sketch of how to generalize the result for a decomposition admitting only partial sets. The key is in relaxing the definition of a trap. Under the new scheme, for every $i$, for a word $\epsilon/2$-far from $P$ (rather than $P_i$), the "partial" trap $\mathcal{Q}_i$ would be required to contain a witness against $P_i$. The arguments translate almost verbatim to this setting, only one must be careful with the definitions such as Definition 6.12 – the exact wording about the (partial) trap containing a witness against the words under consideration becomes even more important.

# 7 Properties with proximity oblivious $q$-tests

For properties with (1-sided, non-adaptive) proximity oblivious $q$-tests we currently do not know how to provide a universal test with $O(n^{1-\gamma})$ many queries, where $\gamma$ depends only on $q$. Here we present such a universal test where $\gamma$ depends on both $q$ and $\rho(\epsilon/2)$ (and $\epsilon$).

In the following we assume knowledge of the definitions and methods of Section 6. We also assume everywhere that $n$ is large enough for the arguments to follow.

**Definition 7.1.** *For $\gamma \in (0,1)$, the* $\gamma$-universal sampler *selects a set $R \subseteq [n]$ where, for every $i \in [n]$, $\Pr[i \in R] = n^{-\gamma}$.*

*Let $P_1, P_2, \ldots, P_\ell \subseteq \{0,1\}^n$ be properties, each having an oblivious one-sided error $q$-tester with the same detection function $\rho(\epsilon)$. Given oracle access to input $w \in \{0,1\}^n$, the* $\gamma$-universal algorithm *for $\bigcup_{i=1}^{\ell} P_i$ selects a set $R \subseteq [n]$ that is the union of $2\log(\ell)$ sets, each chosen according to the $\gamma$-universal sampler. If $|R| > 4\log(\ell) n^{1-\gamma}$ then it accepts immediately, and otherwise it queries the input on all indexes of $R$, rejects if $R$ is a $P_i$-witness against $w$ for every $i \in [\ell]$, and accepts otherwise.*

The main result that we prove here is the following:

**Theorem 7.2.** *For every property $P$ with a proximity oblivious $q$-test with detection function $\rho(\epsilon)$ there exists $\gamma$ depending on $q$ and $\rho(\epsilon/2)$ (for every $\epsilon$), so that for $n$ large enough and every $\epsilon$-far input over $\{0,1\}^n$, the $\gamma$-universal sampler finds a witness against it with probability $1-o(1)$. Therefor (by amplification and union bound) the $\gamma$-universal test is indeed an $\epsilon$-test for $\bigcup_{i=1}^{\ell} P_i$ for large enough $n$.*

We will make crucial use of sunflowers.

**Definition 7.3.** *A* sunflower with center *$A$ is a family of subsets $B_1,\ldots,B_t \subseteq \{1,\ldots,n\}$ so that every $B_i$ contains $A$, and $B_1,\ldots,B_t$ are disjoint outside of $A$ (a completely disjoint family is a sunflower with center $A=\emptyset$).*

**Lemma 7.4** (sunflower theorem, Erdös and Rado [9])**.** *Any family of at least $s = q!t^{q+1}$ sets whose sizes are at most $q$ contains a sub-family of size $t$ which is a sunflower.*

In the following $q$ would be the (constant) number of queries of the proximity-oblivious test, and $t$ would be some power of $n$, so our required $s$ would essentially be another power of $n$.

We next define fragments.

**Definition 7.5** (fragments and violations)**.** *A fragment $\xi = (A, v)$ consists of a subset $A \subseteq \{1,\ldots,n\}$ and a function $v : A \to \{0,1\}$. The special case where $A = \emptyset$ is called the* null fragment*.*

*A fragment $\xi_1 = (A_1, v_1)$* contains *$\xi_2 = (A_2, v_2)$ if $A_2 \subseteq A_1$ and the restriction of $v_1$ to $A_2$ is $v_2$; in this case the* difference fragment *$\xi_3 = (A_3, v_3) = \xi_1 \setminus \xi_2$ is defined where $A_3 = A_1 \setminus A_2$ and $v_3$ is the restriction of $v_1$ to $A_3$.*

*A fragment $\xi = (A, v)$ is said to be* violated *by the input $w$ if the restriction of $w$ to $A$ is $v$.*

It will be easier for us to redefine proximity oblivious tests as distributions over fragments.

**Definition 7.6** (fragment version of a $q$-test)**.** *A* proximity oblivious $q$-test *for $P$ is a distribution $\mu$ over a set $\Xi$ of fragments of sizes bounded by $q$ (some members of $\Xi$ could be with probability $0$) satisfying the following:*

- *If $w$ satisfies $P$ then no fragment is violated (not even probability $0$ ones).*
- *If $w$ is $\epsilon$-far from $P$ then the probability of picking a violated fragment is at least $\rho(\epsilon)$.*

When moving from a $q$-test as in the original definition of finding a witness against $w$, to a $q$-test as per the above definition, the original $\rho(\epsilon)$ might be divided by up to $2^q$ (every original query set is converted to all corresponding fragments that are possible witnesses against the input).

We next define how fragments can be "shortened" sometimes, through either queries or logic.

**Definition 7.7** (witnesses and refutations)**.** *A* witness *for a fragment $\xi$ is a containing fragment $\xi'$, so that the difference $\xi' \setminus \xi$ is violated by the input $w$ ($\xi$ itself does not have to be violated by $w$).*

*A* refutation *for a fragment $\xi$ is a set $\Xi$ of fragments, one of which containing $\xi$, so that no possible input that satisfies the entire set $\Xi$ may satisfy $\xi$.*

Note that in particular a set $\Xi$ is a refutation of the null fragment if and only if it is unsatisfiable.

Our main tool of analyzing the universal sampler is the following:

**Definition 7.8** ($R$-reduction of a test)**.** *Given a $q$-test for a property $P$, as a distribution $\mu$ over a set $\Xi$ of fragments, and a set of queries $R \subseteq \{1, \ldots, n\}$, the* $R$-reduction *of the test is the result of the following process.*

1. *For every $i \in R$, we add the corresponding violated fragment $(i, \neg w(i))$, where $w$ is the input, to $\Xi$, for the time being with probability 0 (this is essentially "adding the query $i$").*
2. *We add to $\Xi$ (still with probability 0) every fragment for which there is a refutation in $\Xi$ (note that, because of the previous item, this also includes fragments for which there is a witness whose corresponding difference was indeed verified to violate $w$ through $R$).*
3. *For every fragment $\xi \in \Xi$ which contains another fragment in $\xi' \in \Xi$ (and is hence made "redundant" by it), we remove $\xi$ from $\Xi$. If $\mu(\xi)$ was non-zero, we modify $\mu$ by adding this probability to the contained $\xi'$ (we can pick any arbitrary $\xi'$ which is not in itself contained in yet another member of $\Xi$).*

The $R$-reduction in itself is not necessarily a test for $P$. It may reject members of $P$, and it may even contain the null fragment (when that happens $\Xi$ will contain only the null fragment and with probability 1; this in particular means that $R$ is a witness for the property against the input, i.e., the input's restriction to $R$ is not extensible to any possible string satisfying $P$).

On the other hand, the following is immediate.

**Observation 7.9.** *For every possible $R$, the $R$-reduction of the test is still a probability distribution over fragments. Moreover, for every possible input $w$, the probability of rejection (obtaining a violating fragment) by the $R$-reduction is at least the corresponding probability by the original test.*

Our main argument for Theorem 1.11 lies in the following: We prove that certain events concerning $R$ and the resulting $R$-reduction of the test occur with probability $1 - o(1)$. Given those events, we prove that if the null fragment is not in the resulting $\Xi$, then it may not be the case that all $\epsilon/2$-far inputs are rejected with probability $\rho(\epsilon/2)$ by the original test (we will construct too many "disjoint" inputs).

In the following, $\gamma$ (of the universal sampler) will be chosen small enough as a function of all other parameters that will be defined. In the following we also view the universal sampler as consisting of $q$ rounds, where in every round, every index is chosen with probability $n^{-2\gamma}$ ($n$ is assumed large enough for this assumption to be viable).

First we define the following with respect to a $\beta$ that will be chosen later ($\gamma$ will depend on $\beta$).

**Definition 7.10** (sunflowers of fragments and fragment generations)**.** *The family of fragments $\xi_1 = (A_1, v_1), \ldots, \xi_t = (A_t, v_t)$ is called a* sunflower with center $\xi = (A, v)$ *if $A_1, \ldots, A_t$ is a sunflower (of sets) with center $A$, and additionally the restriction of every $v_i$ to $A$ is $v$.*

*Given a $q$-test with the set of fragments $\Xi$, all members of $\Xi$ are said to be* generation 0*. By induction, a fragment is said to be* generation $i$ *if it is the center of a sunflower of $n^\beta$ fragments whose generation is at most $i - 1$ and which are all witnesses for it, or it has a refutation using fragments whose generation is at most $i$ (and unless the fragment is already of a smaller generation).*

*Fragments which do not have a designated generation by the above are said to be* generation $\infty$*.*

We will only be interested in fragments of generation up to $q$ due to this simple observation.

**Observation 7.11.** *A generation $i$ fragment for $i < \infty$ has length at most $q - i$, so in particular all finite generations are at most $q$.*

A central claim is the following:

**Lemma 7.12.** *Let $R = \bigcup_{j=1}^{q} R_j$ be the result of $q$ rounds where in each round every index $i$ is independently chosen to be in $R_j$ with probability $n^{-2\gamma}$. With probability $1 - o(1)$, after the $j$'th round, the $\bigcup_{k=1}^{j} R_k$-reduction of the test contains all generation $j$ fragments or sub-fragments thereof. This is when $\gamma$ is chosen to be $\beta/(4q)$.*

*Proof.* This is proved by induction. The base is $j = 0$ (the $\emptyset$-reduction of the test will still have all the original fragments, or sub-fragments thereof if there were meaningful refutations).

Let us assume that the $\bigcup_{k=1}^{j-1} R_k$-reduction of the test includes all generation $j-1$ fragments or sub-fragments thereof. For a generation $j$ fragment $\xi = (A, v)$ that is the center of a sunflower of witnesses, first let $\xi' = (A', v')$ be any member thereof. The probability that $R_j$ contains $A' \setminus A$ is at least $n^{-2q\gamma}$. Note that when this happens, $\xi$ or a sub-fragment thereof will be in the $\bigcup_{k=1}^{j} R_k$-reduction as required.

Now there are at least $n^{\beta}$ members of the sunflower, and the events of each difference to be included in $R_j$ are all independent (as this is a sunflower). Therefor the probability of none of the events happening is at most $(1 - n^{-2q\gamma})^{n^{\beta}} < \exp(-n^{\beta - 2q\gamma})$, which is $o(n^{-1-q})$ taking $\gamma = \beta/(4q)$. Noting that there are not more than $n^{1+q}$ fragments in all, we are done for all such flower centers by a union bound.

The case where the generation $j$ fragment has a refutation by other generation $j$ fragments is immediate, once we know that all fragments that are generation $j$ through being a center of a sunflower are included. □

This claim in turns motivates the following definition.

**Definition 7.13.** *Given a test (as a distribution over a set of fragments $\Xi$) and an input $w$, the* generational reduction *thereof is the result of the following process:*

1. *We add to $\Xi$ (with probability 0) every fragment which is of generation $i$ for some $i < \infty$ (and hence $i \leq q$).*
2. *For every fragment $\xi \in \Xi$ which contains another fragment in $\xi' \in \Xi$ (and is hence made "redundant" by it), we remove $\xi$ from $\Xi$. If $\mu(\xi)$ was non-zero, we modify $\mu$ by adding this probability to the contained $\xi'$ (we can pick any arbitrary $\xi'$ which is not in itself contained in yet another member of $\Xi$).*

Again the following is straightforward.

**Observation 7.14.** *The generational reduction of the test is still a probability distribution over fragments. Moreover, for every possible input $w$, the probability of rejection (obtaining a violating fragment) by the generational reduction is at least the corresponding probability by the original test.*

It is important for us to note the following, as the generational reduction has a better structure than just any randomized $R$-reduction obtained through sampling.

**Lemma 7.15.** *With probability $1 - o(1)$, the $R$-reduction of the test is also a reduction of the generational reduction of the test.*

*Proof.* This is equivalent to Lemma 7.12 for $j = q$, because it means that with probability $1 - o(1)$ there will be witnesses in $R$ to all finite generation fragments, recalling also Observation 7.11. □

In particular, if the generational reduction has the null fragment in its set of fragments, then with probability $1-o(1)$ the $\gamma$-universal testing algorithm will reject the property. To complete the components required for the proof of Theorem 1.11, we will assume that the null fragment is not in this reduction (i.e. it is of generation $\infty$, which is equivalent to $R$ not being a witness against the input) and reach a contradiction. At this point we use the sunflower theorem.

**Lemma 7.16.** *Let $\Xi_R$ denote the set of fragments of the generational reduction of the test, and assume that it does not contain the null fragment. For $n$ large enough, there exists no fragment (regardless of whether it is violated itself) that is contained in more than $n^{(q+2)\beta}$ members of $\Xi_R$ that are witnesses against it.*

*Proof.* If $\xi = (A, v)$ was such a fragment, and $\xi_1 = (A_1, v_1), \ldots, \xi_t = (A_t, v_t)$ were (containing) members of $\Xi_R$ for $t = n^{(q+2)\beta}$ that witness it, then (for $n$ such that $n^\beta > q!$) by Lemma 7.4 there would have been a sunflower of sets $A_{j_1}, \ldots, A_{j_t}$ for $t = n^\beta$, whose center is some set $A'$ that contains $A$. Now the restrictions of $v_{j_1}, \ldots, v_{j_t}$ to $A'$ are all identical: Over $A$ these are identical to $v$, and over $A' \setminus A$ these are identical to the restriction of $w$ to this set. Let $v'$ denote the common restriction of $v_{j_1}, \ldots, v_{j_t}$ to $A'$. $\xi_{j_1}, \ldots, \xi_{j_t}$ are now also a sunflower of fragments, all witnesses to their center $\xi' = (A', v')$. This would have meant that $\xi'$ is a fragment of some finite generation, which is a contradiction to $\Xi_R$ already corresponding to the generational reduction of the test. □

In particular (through the null fragment), the above means that there are no more than $n^{(q+2)\beta}$ members of $\Xi_R$ that are violated by $w$. However, $\Xi_R$ in itself could still be very large, for example it could contain many fragments that would be violated by the bit-wise negation of $w$.

In the following, we assume that $w$ is an $\epsilon$-far word for which the generational reduction does not contain the null fragment. We then do the following construction.

**Definition 7.17.** *Assume that $\Xi_R$ does not contain the null fragment (and is hence satisfiable). We define by induction the following sequences, where $w_0 = w$, $\Xi_0 = 0$ and $B_0 = \emptyset$. We let $w^*$ be any word that violates no member of $\Xi_R$.*

- *$\Xi_i$ is the set of the members of $\Xi_R$ that are violated by $w_{i-1}$.*
- *$B_i = B_{i-1} \cup \bigcup\{A : \xi = (A, v) \in \Xi_i\}$.*
- *$w_i$ is identical to $w^*$ over $B_i$ and identical to $w$ outside of it.*

We are interested in $w_0, \ldots, w_r$, $\Xi_1, \ldots, \Xi_{r+1}$ and $B_1, \ldots, B_r$ for $r = 1/\rho(\epsilon/2)$. The following lemma gives us their required properties.

**Lemma 7.18.** *Assume that $R$ is not a witness, and equivalently $\Xi_R$ does not contain the null fragment. All of the following hold for $n$ large enough.*

- *The sets $\Xi_i$ are all disjoint.*
- *$|\Xi_1| \leq n^{(q+2)\beta}$ and $|B_1| \leq qn^{(q+2)\beta}$.*
- *$|\Xi_i| \leq n^{(q+2)\beta}|B_{i-1}|^q$ and $|B_i| \leq |B_{i-1}| + qn^{(q+2)\beta}|B_{i-1}|^q$ for $i > 1$.*
- *$|B_k| \leq n^{(5q)^k\beta}$ for $k > 1$.*

*Proof.* The first item is because $\Xi_i$ cannot contain any fragment whose respective set is inside $B_{i-1}$ (because $w_{i-1}$ is identical to $w^*$ there), or any fragment whose respective set is not contained in $B_i$ (because of how $B_i$ was defined), and we have successive containment $B_{i-1} \subseteq B_i$.

The second item is from the discussion after Lemma 7.16, noting also that all members of $\Xi_R$ are of length bounded by $q$.

The third item is by Lemma 7.16 again. We note that violated fragments of $\Xi_R$ here can only come from witnesses in $\Xi_R$ for fragments inside $B_{i-1}$, and there are less than $|B_{i-1}|^q$ relevant fragments (all possible restrictions of $w^*$ to subsets of size at most $q$ of $B_{i-1}$).

The fourth item is by basic numeric induction. □

Now we finally have all the components for proving Theorem 1.11.

*Proof of Theorem 1.11.* We take $\beta = \frac{1}{2}(5q)^{-r}$ (where $r = 1/\rho(\epsilon/2)$ and $n$ to be large enough so that $n^{-1/2} \leq \epsilon/2$). By Lemma 7.15, with probability $1-o(1)$ the $R$-reduction of the test will include also its generational reduction. To conclude we prove that such an $R$ is necessarily a witness against the input. Let us assume on the contrary that the $R$-reduction (and hence the generational reduction) does not contain the null fragment.

We refer to the construction of Definition 7.17. The choice of parameters above and Lemma 7.18 ensure that $|B_r| \leq \epsilon n/2$, and so all the inputs $w_0, \ldots, w_r$ are $\epsilon/2$-close to $w$ and hence are $\epsilon/2$-far from the property. Hence the original $q$-test and its generational reduction have to reject each of those inputs with probability at least $\rho(\epsilon/2)$. However, this means that in $\Xi_R$ there are $r+1$ disjoint subsets $\Xi_1, \ldots, \Xi_{r+1}$ where each of which is assigned a probability of at least $\rho(\epsilon/2)$ by the generational reduction, which is the contradiction to the sum of all probabilities being 1. □

# 8 Highly decomposable properties

We prove here Theorem 1.13. The property that we will pick to show it will be the following one of being $k$-paritic.

**Definition 8.1.** *A binary string $w = (w_1, \ldots, w_n) \in \{0,1\}^n$ is called* k-paritic *if there exist $i_1, \ldots, i_k$ for which $i_1 = 1$, $i_j + n/2k \leq i_{j+1}$ for all $1 \leq j < k$ and $i_k + n/2k \leq n$, such that for every $0 \leq r < n/2k$ we have $\bigoplus_{j=1}^{k} w_{i_j+r} = 0$.*

*For fixed $i_1, \ldots, i_k$ as above, we let $P_{i_1,\ldots,i_k}$ denote the property of satisfying $\bigoplus_{j=1}^{k} w_{i_j+r} = 0$ for every $0 \leq r < n/2k$ (for these particular $i_1, \ldots, i_k$).*

Theorem 1.13 then immediately follows from Lemma 8.2 and Lemma 8.5 below.

**Lemma 8.2.** *The property of being $k$-paritic is decomposable to at most $n^{k-1}$ many properties, so that each of them has a proximity-oblivious 1-sided $k$-test with detection function $\rho(\epsilon) = O(k\epsilon)$.*

*Proof.* We decompose the property of being $k$-paritic to the properties $P_{i_1,\ldots,i_k}$ (as in Definition 8.1) where $i_1, \ldots, i_k$ are any indexes such that $i_1 = 1$, $i_j + n/2k \leq i_{j+1}$ for all $1 \leq j < k$ and $i_k + n/2k \leq n$ (note that these properties need not be disjoint). There are less than $n^{k-1}$ such properties ($i_1$ has one value and every other $i_j$ clearly can have less than $n$ possible values), and clearly their union is the property of being $k$-paritic.

The proximity-oblivious $k$-test for every property $P_{i_1,\ldots,i_k}$ is done by taking a uniformly drawn value from $\{0, \ldots, n/2k\}$ for $r$, and checking that the parity requirement $\bigoplus_{j=1}^{k} w_{i_j+r} = 0$ is satisfied

(which uses $k$ queries). To get at the $O(k\epsilon)$ bound on the detection function, we note that for $w$ to be $\epsilon$-far from $P_{i_1,\ldots,i_k}$, at least $\epsilon n$ values of the possible $n/2k$ values for $r$ must be such that $\bigoplus_{j=1}^{k} w_{i_j+r} = 1$, so the probability to get such a value for $r$ is $\frac{\epsilon n}{n/2k} = O(k\epsilon)$. □

Before continuing we show that being $k$-paritic is not too dense.

**Lemma 8.3.** *For every fixed $k$, a uniformly random member of $0,1^n$ (each bit being chosen uniformly and independently) is $1/10k$-far from being $k$-paritic with probability $1 - o(1)$.*

*Proof.* First we consider a property $P_{i_1,\ldots,i_k}$ for specific $i_1, \ldots, i_k$ as in Definition 8.1. For every $0 \leq r < n/2k$, the probability for $\bigoplus_{j=1}^{k} w_{i_j+r} = 1$ is exactly $\frac{1}{2}$, and these events are completely independent for different values of $r$. Hence by a straightforward large deviation inequality with probability at least $1-2^{n/10k}$ it holds that we have a set $R \subset \{0, \ldots, \lceil n/2k \rceil - 1\}$ of size at least $n/10k$ so that for every $r \in R$ we have $\bigoplus_{j=1}^{k} w_{i_j+r} = 1$. When this occurs the word $w = (w_1, \ldots, w_n)$ is clearly $1/10k$-far from $P_{i_1,\ldots,i_k}$.

The lemma now follows from a union bound over all properties $P_{i_1,\ldots,i_k}$ (whose number is less than $n^{k-1}$, see the proof of Lemma 8.2). □

We will also use a traditional Yao's argument. The following is similar to the form that appears in [10] (but was developed earlier).

**Lemma 8.4.** *Suppose that $\mathcal{D}_P$ and $\mathcal{D}_N$ are two distributions over $\{0,1\}$. For an index set $Q \subset \{1, \ldots, n\}$ of size $q$ a word $v \in \{0,1\}^q$, let $\alpha(Q, v)$ be the probability that a word $w \in \{0.1\}$ drawn according to $\mathcal{D}_P$ agrees with $v$ over $Q$ (i.e., that setting $i_1, \ldots, i_q$ to be the members of $Q$ in sorted order, we have $w_{i_j} = v_j$ for all $1 \leq j \leq q$). Define $\beta(Q, v)$ similarly with $\mathcal{D}_N$ instead of $\mathcal{D}_P$.*

*If for every $Q$ of size $q$ and every $v \in \{0,1\}^q$ we have that $\alpha(Q, v) \leq (1 - \eta)\beta(Q, v)$, then no algorithm making up to $q$ queries can distinguish with probability more than $\eta$ (even an adaptive one and in a 2-sided manner) between the case where $w$ was drawn according to $\mathcal{D}_P$ and the case where it was drawn according to $\mathcal{D}_N$.*

The following now concludes the proof of Theorem 1.13

**Lemma 8.5.** *For any fixed $k$, The property of being $k$-paritic in itself cannot be $1/5k$-tested using $o(n^{1-1/k})$ many queries (even not by 2-sided adaptive algorithms).*

*Proof.* Here we use Yao's method as per outlined in [10] (but developed earlier). We will assume that $n = 2kl$ for some integer $l$, as the move from this to general $n$ involves simple padding. We define two distributions.

- The distribution $\mathcal{D}_P$ starts by first choosing uniformly and independently $2l(j-1) + 1 \leq i_j \leq 2l(j-1) + l$ for every $1 \leq j \leq k$. Then we take $w \in \{0,1\}^n$ to be a uniformly random member of $P_{i_1,\ldots,i_k}$, the corresponding property defined in the proof of Lemma 8.2 (out of the $2^{n-l}$ members thereof).
- The distribution $\mathcal{D}_N$ is just the uniform distribution over $\{0,1\}^n$.

It is clear that an input drawn according to $\mathcal{D}_P$ is always $k$-paritic. Also, by Lemma 8.3 we have that with probability $1 - o(1)$ an input drawn according to $\mathcal{D}_N$ is $1/5k$-far from being $k$-paritic.

Also note that for every $v \in \{0,1\}^q$ and an index set $Q \subseteq \{1, \ldots, n\}$ of size $q$, the probability of a word $w$ drawn according to $\mathcal{D}_N$ to agree with $v$ over $Q$ is exactly $2^{-q}$. To complete the argument,

by Lemma 8.4 it remains to show that for every $v \in \{0,1\}^q$ and every $Q \subseteq \{1,\ldots,n\}$ of size $q$ where $q = o(n^{1-1/k})$, the probability for such an agreement is at least $(1-o(1))2^{-q}$.

Let $E$ be the event that there is no $0 \leq r < l$ for which $\{j_1 + r, \ldots, j_k + r\} \subset Q$. Conditioned on $E$, the probability of $w$ to agree with $v$ over $Q$ is exactly $2^{-q}$, so it remains to show that $E$ occurs with probability $1 - o(1)$. Let $s_1, \ldots, s_k$ be members of $Q$. The only case where there can be a positive probability for the equalities $i_1 + r = s_1, \ldots, i_k + r = s_k$ is if $2l(j-1)+1 \leq s_j \leq 2lj$ for every $1 \leq j \leq k$, and also in this case the probability for all equalities to occur is bounded by $l^{1-k} = (2k/n)^{k-1}$ (by a union bound argument over the $l$ possible values for $r$).

The number of possible eligible $k$-tuples $s_1, \ldots, s_k$ in $Q$ is at most $(q/k)^k$. By the union bound the probability for $E$ not to occur is then bounded by $(2q/n)^{k-1}(q/k)$. For a fixed $k$, if $q = o(n^{1-1/k})$ then this probability bound evaluates to $o(1)$, concluding the proof. □

It would be interesting to find out whether there exists a property decomposable into a relatively small number of testable properties that in itself requires a linear number of queries to test. The following standard proposition shows that for being $k$-paritic our lower bounds are about as far they go.

**Proposition 8.6.** *The property of being $k$-paritic is testable by a non-adaptive 1-sided test, making $O(n^{1-1/k}(\log(n)/\epsilon)^{1/k})$ queries and detecting $\epsilon$-far inputs with constant probability,*

*Proof.* We will use the following algorithm:

- Choose a query set $Q$ by choosing for every $1 \leq j \leq n$ independently whether $j \in Q$, where this occurs with probability $(10kn^{-1}\log(n)/\epsilon)^{1/k}$.
- If $|Q| > 2n(10kn^{-1}\log(n)/\epsilon)^{1/k}$ then accept the input without making any queries (by a large deviation inequality this occurs with probability $o(1)$).
- Otherwise, make all the queries in $Q$, and accept the input if and only if there exists a $k$-paritic word $u \in \{0,1\}^n$ whose restriction to $Q$ agrees with all queries made to the input word $w$.

Clearly, if the input word $w$ is $k$-paritic then it will always be accepted, either arbitrarily in the second step or by $u = w$ in the third step. It now remains to prove that $\epsilon$-far words are rejected with high probability. The second step assures that the number of queries is always at most $2n(10kn^{-1}\log(n)/\epsilon)^{1/k} = O(n^{1-1/k}(\log(n)/\epsilon)^{1/k})$ (rather than being so only with probability $1-o(1)$).

We may safely ignore the case where there is acceptance in the second step as it occurs with probability $o(1)$, and henceforth analyze the algorithm as if this step was removed from it. We start by analyzing the property $P_{i_1,\ldots,i_k}$ for specific $i_1, \ldots, i_k$ as in Definition 8.1. If $w$ is $\epsilon$-far from $P_{i_1,\ldots,i_k}$, then there is a set $R \subset \{0,\ldots,\lceil n/2k \rceil - 1\}$ of size at least $\epsilon n$ so that for every $r \in R$ we have $\bigoplus_{j=1}^{k} w_{i_j+r} = 1$. For every fixed $r \in R$, the probability to query its corresponding witness of not being in $P_{i_1,\ldots,i_k}$, i.e. the probability for $\{i_1 + r, \ldots, i_k + r\} \in Q$, is $10kn^{-1}\log(n)/\epsilon$.

The above means that for the specific property $P_{i_1,\ldots,i_k}$, the probability of not detecting a witness for the input not being in it is at most $(1 - 10kn^{-1}\log(n)/\epsilon)^{\epsilon n} < \exp(-10k\log(n)) = o(n^{k-1})$. All that remains to do is to perform a union bound over all properties $P_{i_1,\ldots,i_k}$, whose union is the property of being $k$-paritic (see Lemma 8.2 and its proof), to see that with probability $1-o(1)$ our query set is such that there is no $k$-paritic word $u$ whose restriction to $Q$ agrees with the queries made to $w$. □

## References


[1] Noga Alon, László Babai, and Alon Itai. A fast and simple randomized parallel algorithm for the maximal independent set problem. *Journal of Algorithms*, 7(4):567–583, December 1986.

[2] Noga Alon, Eldar Fischer, Michael Krivelevich, and Mario Szegedy. Efficient testing of large graphs. *Combinatorica*, 20(4):451–476, 2000.

[3] Noga Alon, Michael Krivelevich, Ilan Newman, and Mario Szegedy. Regular languages are testable with a constant number of queries. *SIAM J. Comput*, 30(6):1842–1862, 2000.

[4] Noga Alon and Joel Spencer. *The Probabilistic Method*. John Wiley & Sons, 2011.

[5] Eli Ben-Sasson, Oded Goldreich, Prahladh Harsha, Madhu Sudan, and Salil P. Vadhan. Robust PCPs of proximity, shorter PCPs, and applications to coding. *SIAM J. Comput*, 36(4):889–974, 2006.

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

[7] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. *J. Comput. Syst. Sci.*, 47(3):549–595, 1993.

[8] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, New York, USA, 2000.

[9] Paul Erdős and Richard Rado. Intersection theorems for systems of sets. *J. London Math. Soc.*, 35:85–90, 1960.

[10] Eldar Fischer. The art of uninformed decisions: A primer to property testing. *Current Trends in Theoretical Computer Science: The Challenge of the New Century*, I:229–264, 2004.

[11] Eldar Fischer and Arie Matsliah. Testing graph isomorphism. *SIAM J. Comput*, 38(1):207–225, 2008.

[12] E. N. Gilbert. A comparison of signalling alphabets. *The Bell System Technical Journal*, 31(3):504–522, May 1952.

[13] Oded Goldreich, Shaffi Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *J. ACM*, 45:653–750, July 1998.

[14] Oded Goldreich and Dana Ron. On sample-based testers. *Electronic Colloquium on Computational Complexity (ECCC)*, (109), 2013.

[15] Tom Gur and Ron Rothblum. Non-interactive proofs of proximity. *Electronic Colloquium on Computational Complexity (ECCC)*, (078), 2013.

[16] J. Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48(4):798–859, 2001.

[17] Solomon Kullback and Richard A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, March 1951.

[18] F. J. MacWilliams and N. J. A. Sloane. *The Theory of Error-Correcting Codes*, volume 16 of *North-Holland Mathematical Library*. North-Holland, 1977.

[19] F. J. MacWilliams, N. J. A. Sloane, and J. G. Thompson. Good self dual codes exist. *Discrete Math.*, 3:153–162, 1972.

[20] Mark S. Pinsker. *Information and information stability of random variables and processes*. Holden-Day Inc., San Francisco, Calif., 1964. Translated and edited by Amiel Feinstein.

[21] Mark D. Reid and Robert C. Williamson. Generalised pinsker inequalities. In *COLT*, 2009.

[22] Dana Ron. Property testing: A learning theory perspective. *Found. Trends Mach. Learn.*, 1:307–402, March 2008.

[23] Dana Ron. Algorithmic and analysis techniques in property testing. *Foundations and Trends in Theoretical Computer Science*, 5(2):73–205, 2010.

[24] Guy N. Rothblum, Salil Vadhan, and Avi Wigderson. Interactive proofs of proximity: delegating computation in sublinear time. In *Proceedings of the 45th annual ACM symposium on Symposium on theory of computing*, STOC '13, pages 793–802, New York, NY, USA, 2013. ACM.

[25] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM J. Comput.*, 25(2):252–271, 1996.

[26] R. R. Varshamov. Estimate of the number of signals in error correcting codes. In *Dokl. Akad. Nauk SSSR*, volume 117, pages 739–741, 1957.