Electronic Colloquium on Computational Complexity, Revision 4 of Report No. 73 (2021)

New sampling lower bounds via the separator*

Emanuele Viola†

August 4, 2023

Abstract

Suppose that a target distribution can be approximately sampled by a low-depth decision tree, or more generally by an efficient cell-probe algorithm. It is shown to be possible to restrict the input to the sampler so that its output distribution is still not too far from the target distribution, and at the same time many output coordinates are almost pairwise independent.

This new tool is then used to obtain several new sampling lower bounds and separations, including a separation between AC0 and low-depth decision trees, and a hierarchy theorem for sampling. It is also used to obtain a new proof of the Patrascu-Viola data-structure lower bound for Rank, thereby unifying sampling and data-structure lower bounds.

*A preliminary version appeared as [Vio23]. This version contains a new result, in Section 9.

†Supported by NSF CCF awards 1813930 and 2114116.

ISSN 1433-8092

.

1 Introduction and our results

Obtaining *computational lower bounds* is a fundamental agenda in theoretical computer science, see for example the textbooks [Juk12, AB09]. One of the most famous lower bounds is the AC0 lower bound for computing the parity function, which separates small AC0 circuits from models that can compute parity. Another direction that has received much attention is the relationship between AC0 and *low-depth decision trees*. The simple Or function on n bits requires decision trees of depth n to be computed exactly, but the picture is more subtle and useful when we consider *average-case* computation, that is we allow errors on a small fraction of inputs. Indeed, *switching lemmas* [FSS84, Ajt83, Yao85, Hås87, SBI04, Raz15, BIS12, IMP12, Hås14] (see [Vio22] for an exposition and discussion) can be interpreted as non-trivial simulations of small AC0 circuits by decision trees. On the other hand, functions such as *Tribes* (see, e.g., [O'D14]), computable by a polynomial-size DNF circuits, require large-depth decision trees, even on average.

In this work we study lower bounds and separations in the setting of *sampling*. This is a challenging generalization of average-case complexity, where we seek to bound the resources required to sample approximately a target distribution, given random bits. The study of *sampling lower bounds* [Vio12b, LV12, Vio14, DW11, BIL12a, BCS14, Vio12c, Vio20, CGZ21, FLRS23] has seen significant activity and progress in the last ten years; for a survey talk see [Vio]. This study has also had impact on other areas. For example, it has had an impact on breakthrough constructions of *two-source extractors*: the papers [CZ16, Li16, CS16, Coh16, BDT16] build on models or results from the study of sampling lower bounds. Also, sampling lower bounds have been used to obtain *data-structure* lower bounds [Vio12b]. In fact, jumping ahead, this paper will further develop this connection to data structures.

Sampling lower bounds for AC0, roughly corresponding to the classical result mentioned above that Parity is not in AC0, have been obtained in [LV12, Vio14, BIL12a, Vio20]. All these lower bounds share common techniques. Interestingly, essentially no technique was known to obtain separations within AC0. The main goal and motivation for this paper is thus to develop new techniques for sampling lower bounds, and apply them to obtain separations within AC0, in particular separating decision-tree from AC0 samplers and obtaining a hierarchy theorem (see Corollary 9). In addition, the new technique is used to "unify" sampling and data-structure lower bounds, that is, to obtain data-structure lower bounds as a consequence of sufficiently strong sampling lower bounds.

The model. The main computational model in this work is a generalization of the decision-tree model known as the *cell-probe* model [Yao81]. Here the input is divided into *words* (a.k.a. *cells)* of w bits, the output is a tuple of *queries*, and each query can be computed by making q probes into the input, adaptively. This model is extensively studied in algorithms, where w corresponds to the *register size* and q to *time*. We note that for $w = 1$ each output query is computed by a *decision tree* of depth q. For larger w each query is also computed by a tree but each internal node probes a word and has 2^w children. Because the output is a tuple of queries, we have several trees of depth q, one for each output query, and so we refer to the algorithm as to a depth-q *forest*. We place no restriction on the number of input words, which we indicate with $\mathbb{N}$. But for concreteness one can replace $\mathbb{N}$ with any large enough integer – as we do later in the proofs. We summarize the model and its key parameters:

Definition 1. We say that $f : W^{\mathbb{N}} \to \Sigma^m$ is a depth-q *forest* with word size w and output alphabet Σ if $W = \{0,1\}^w$ and $f = (f_1, f_2, \ldots, f_m)$ where each $f_i : W^{\mathbb{N}} \to \Sigma$ is a depth-q decision tree where the variables are over W, each internal node has $|W|$ children, and the leaves are labeled with elements from Σ.

The main goal of this paper is to show that a target distribution S over Σ^m is hard to sample by a low-depth forest. We measure the distance between distributions X and Y over D using *statistical (*a.k.a. *total variation, L_1) distance*

$$\Delta(X,Y) := \max_{T \subseteq D} |\mathbb{P}[X \in T] - \mathbb{P}[Y \in T]|.$$

So the lower-bound goal is to show $\Delta(f(U_{W^{\mathbb{N}}}), S)$ is large for any low-depth forest f, where $U_{W^{\mathbb{N}}}$ is the uniform distribution over $W^{\mathbb{N}}$. In general for a set H we write U_H for the uniform distribution over H, and simply U when H is clear from the context.

Previous sampling lower bounds. Before this work, essentially the only sampling lower bounds in the cell-probe model, or even in the decision-tree model, were those that followed from the sampling lower bounds for AC0 circuits [LV12, Vio14, BIL12a, Vio20] – using the fact that a depth-q tree can be written as a DNF with width qw. This was unsatisfactory for several reasons. First, such results obviously cannot be used to prove separations within AC0. By contrast, this paper obtains such separations. Second, the AC0 lower bounds only hold for sampling *pseudorandom objects*, such as extractors or error-correcting codes. By contrast, the lower bounds in this paper apply to fundamental data-structure problems that do not have pseudorandom properties, and this allows us to unify sampling and data-structure lower bounds.

1.1 Our results

In this work we prove new sampling lower bounds and use them to derive a number of new separations. We emphasize that our results are new even for decision trees, corresponding to word size $w = 1$. However, we obtain stronger results by considering larger word size. Similarly, the lower bounds we prove were not known even for statistical distance $\Omega(1)$. But in fact we prove stronger bounds, where the statistical distance is exponentially close to 1. Via technically simple connections, the first of which was pointed out in [Vio12b], this generality enables several applications discussed below. In particular, jumping ahead, it will allow us to unify sampling and data-structure lower bounds. Indeed, our lower bounds hold for fundamental problems in data structures.

First we obtain a sampling lower bound for the distribution $\text{RANK}(U_{\{0,1\}^m})$ where RANK is defined next.

Definition 2. For $x \in \{0,1\}^m$ define $\text{RANK}(x)$ as the string $y \in \{0,1,\ldots,m\}^m$ where $y_i := \sum_{j \le i} x_j$ is the rank of i.

Example 3. $\text{RANK}(0,1,0,1) = (0,1,1,2)$.

Theorem 4. *Let $f : W^{\mathbb{N}} \to \{0, 1, \ldots, m\}^m$ be a depth-q forest with word size $w \geq \log m$. Then* $\Delta(f(U_{W^{\mathbb{N}}}), \textsc{Rank}(U_{\{0,1\}^m})) \geq 1 - 2 \cdot 2^{-m/w^{O(q)}}$.

Throughout this paper, the notation $O(.)$ and $\Omega(.)$ denotes absolute constants.

It follows from [Yu19], which builds on [Pǎt08], that this bound is tight. In particular, we can sample $\textsc{Rank}(U)$ with depth $q = O(\log m)/\log w$ and constant statistical distance

This result can also be interpreted as a negative result for sampling *random walks on graphs*. Consider the graph G over $\{0, 1, \ldots, m\}$ where the neighbors of Node i are $\{i, i + 1\}$. Note that $\textsc{Rank}(U)$ is the sequence of nodes visited during the walk with edge choices U. Theorem 4 proves a lower bound for sampling this random walk. Note that as the graph is fixed, this result applies even if the algorithm depends on the graph.

Next we consider the *predecessor* problem.

Definition 5. For $x \in \{0, 1\}^m$ define $\textsc{Pred}(x)$ as the string $y \in \{0, 1, \ldots, m\}^m$ where $y_i := \max\{j : j \leq i \text{ and } x_j = 1\}$ is the predecessor of i. (Say $y_i = 0$ if there is no $j \leq i$ with $x_j = 1$.)

Unlike $\textsc{Rank}$, it turns out that $\textsc{Pred}$ can be sampled efficiently. Specifically, there is a depth-$O(1)$ forest sampling $\textsc{Pred}(U)$ with statistical distance $1/\mathrm{poly}(m)$. This is just because the predecessor of i can be computed by inspecting the bit positions from $i - q \log n$ to i of x, except with error probability $1/n^q$.

Loosely inspired by works in data-structure lower bounds [PT06], we prove a lower bound for $\textsc{Pred}(X)$ under a distribution X which is tailored for the applications below. This lower bound is really a lower bound for a "direct-product" version of $\textsc{Pred}$, where r instances have to be solved simultaneously. In fact, the bound holds even for the *colored* version, where items have colors and we just need to return the color of the predecessor. Next we define this problem and state our results for it.

Definition 6. For an $r \times m$ matrix M with entries in $\{-, \hat{\bigcirc}, \hat{\bullet}\}$ we define the $r \times m$ *Colored-Multi-Predecessor* matrix $\mathrm{CM}\textsc{Pred}(M)$ with entries in $\{\hat{\bigcirc}, \hat{\bullet}, \bigcirc, \bullet\}$ as follows. For any i, j we define $\mathrm{CM}\textsc{Pred}(M)_{i,j}$ to be:

$M_{i,j}$ if $M_{i,j} \neq -$,

$\bullet$ if the predecessor of j on row i is $\hat{\bullet}$ (that is, there is $j' < j$ such that $M_{i,j'} = \hat{\bullet}$ and for every k such that $j' < k < j$ we have $M_{i,k} = -$), and

$\bigcirc$ otherwise.

The distribution Π on $r \times m$ matrices is defined as follows, for m divisible by w^r. Divide row $i = 1, 2, \ldots, r$ in consecutive blocks of w^i elements. For each block, pick a uniform element, and assign to it a uniform element from $\{\hat{\bigcirc}, \hat{\bullet}\}$. All the other elements are set to $-$.

Example 7. $\mathrm{CM}\textsc{Pred}\left(\begin{bmatrix} - & \hat{\bullet} & - & \hat{\bigcirc} \\ - & - & \hat{\bullet} & - \end{bmatrix}\right) = \begin{bmatrix} \bigcirc & \hat{\bullet} & \bullet & \hat{\bigcirc} \\ \bigcirc & \bigcirc & \hat{\bullet} & \bullet \end{bmatrix}.$

Working with the alphabet $\{\hat{\bigcirc}, \hat{\bullet}, \bigcirc, \bullet\}$ allows us to reconstruct M from $\mathrm{CM}\textsc{Pred}(M)$, slightly simplifying the argument.

Theorem 8. *There exists a constant c such that for $r = cq$ the following holds.*

Let $f : W^{\mathbb{N}} \to (\{\hat{\bigcirc}, \hat{\bullet}, \bigcirc, \bullet\}^r)^m$ be a depth-q forest with word size $w \geq \log m$.

Let Π be an $r \times m$ random matrix as in Definition 6.

Then $\Delta(f(U_{W^{\mathbb{N}}}), \mathrm{CM}\textsc{Pred}(\Pi)) \geq 1 - 2 \cdot 2^{-m/w^{O(q)}}$.

Motivation for studying CMPred(Π): New separations. The problem $\mathrm{CMP}{\small\mathrm{RED}}(\Pi)$ is designed to be easy to sample with a little more resources than we prove lower bounds for. Thus the theorem gives two separations. First, we obtain a *probe-hierarchy* for sampling: for any q there is an explicit problem that can be sampled exactly with $O(q)$ probes, but only very poorly with q. Second, the same problem can be also sampled by an explicit, polynomial-size DNF, thus giving a separation between sampling by cell probes and DNFs. Such results were not known even for word size $w = 1$, statistical distance 0.01 rather than close to 1, and AC0 instead of DNF.

Proving hierarchies and separations among various restricted computational models is a main research agenda of theoretical computer science. We consider them in the context of sampling. For example, it is a classical result that small DNF circuits can compute functions that require decision trees of large depth, even on average. Our results strengthen this separation substantially.

Corollary 9. *For every m, q, w such that $w \geq \log m$ the following holds.*

There exists a distribution $S \subseteq \left(\{0,1\}^{O(q)}\right)^m$ such that for any depth-q forest f with word size w we have $\Delta(f(U_{W^{\mathbb{N}}}), S) \geq 1 - 2 \cdot 2^{-m/w^{O(q)}}$. But S can be sampled (with distance 0) by both

(1) An explicit depth-$O(q)$ forest with word size w; and

(2) An explicit $\mathrm{poly}(m)$-size DNF.

The distribution in this corollary is $\mathrm{CMP}{\small\mathrm{RED}}(\Pi)$ with $r = O(q)$. To sample it, we can identify row i of Π with a string of $\log_2(2 \cdot w^i)^{m/w^i}$ bits, indicating the choice of color and element ($2 \cdot w^i \leq O(m)$ possibilities) for each of the m/w^i blocks. Color i of query j can be computed from these bits probing $O(1)$ words. Repeating this for $i = 1, 2, \ldots, r$ gives (1) in the corollary. (2) is similar.

Previous attempts to establish a separation between forests and AC0 circuits resulted in (i) Theorem 1.4 in [Vio12b] which applies to *randomness-efficient* samplers, achieves constant statistical distance, and has $w = 1$, and (ii) Theorem 3 in [Vio20] which applies to *non-adaptive* samplers. It is an open question whether $\mathrm{R}{\small\mathrm{ANK}}$ can be sampled by polynomial-size AC0 circuits. Recent results [Yu19] building on [Pǎt08] imply that it can be sampled with $O(\log m)$ probes and constant statistical distance, which gives *quasi*-polynomial size AC0 (by padding this implies a separation between polynomial-size AC0 and forests of depth $\log^c m$ for a constant $c < 1$).

Data structures A *(static) data-structure problem* is a map $f : \{0,1\}^n \to \Sigma^m$, where m queries over alphabet Σ are to be answered about n bits of data. A *data-structure* with word size w for this problem are two functions $g : \{0,1\}^n \to \{0,1\}^{n+r}, h : \{0,1\}^{n+r} \to \Sigma^m$ where g is arbitrary and h is a depth-q forest with word size w such that $f = h \circ g$. That is, we seek to store the n bits of data into $n + r$ bits so that the queries can be computed fast. Note that the $n+r$ bits are divided in words of w bits. We call r the *redundancy* of the data structure, and we focus on the *succinct* regime $r = o(n)$. Many papers are devoted to proving lower bounds in this regime, including [GM07, Vio12a, Gol09, PV10, LY20]; and it is shown in [Vio19] that improving on the long-standing bounds in [GM07] would yield new circuit lower bounds.

The paper [Vio12b] pointed out a technically simple connection between samplers and data-structures: any data structure can be used to sample the distribution $f(U)$ by a depth-q forest with statistical distance $1-2^{-r}$. Simply fill the $n+r$ bits uniformly and run the query algorithms.

A data structure is *equivalent* to the special case of samplers *which just use* $n+r$ *input bits*. But samplers can use *any* number of input bits, and many samplers in the literature do use $(1+\Omega(1))n$ input bits, for example to sample noise vectors, subsets, or permutations, cf. [Vio12b].

Hence, the sampling lower bounds above imply data-structure lower bounds. Theorem 4 gives a new proof of the data-structure lower bound for RANK from [PV10], which was recently shown to be tight in [Yu19], building on [Pǎt08]. This new proof shows that the lower bound applies even to samplers. Informally, this suggests that the "reason" why the lower bound for RANK holds is not that the input is "compressed," *but rather that low-depth forests simply cannot generate the type of dependencies in* RANK, *regardless of their input.*

The program of proving data-structure lower bounds via sampling was suggested a decade ago [Vio12b], but the only previous cell-probe lower bound obtained this way is for error-correcting codes and follows from the AC0 lower bounds [LV12, BIL12b]. This paper shows that this program is feasible for problems such as RANK.

Similarly, we obtain a data-structure lower bound for CMPRED. Also, the sampling hierarchy in Corollary 9 translates to a *data-structure hierarchy*. Hierarchies in data structures have been considered since the 90's. [Mil99] gives a *non-explicit* problem where decreasing the *redundancy* by one bit makes the probe time jump from constant to linear. We give explicit problems where increasing the probe time q by a constant factor makes the redundancy shrink from almost linear to zero. Previous bounds such as [PV10] imply such a result for q about $\log m$. We achieve a broader range including $q = O(1)$. To the best of our knowledge, such a result does not appear in the literature.

Corollary 10. *[Data-structure hierarchy] For every q and m there exists an explicit function $f : \{0,1\}^m \to (\{0,1\}^{O(q)})^m$ which has a data structure with word size w, redundancy zero, and making $O(q)$ probes, but such that any data-structure with word size w making q probes requires redundancy $r \geq m/w^{O(q)}$.*

The sampling viewpoint is not essential for the data-structure lower bound for CMPRED or for Corollary 10: just like RANK, they can be proved without referring to sampling.

Communication protocols. Above we considered one application of proving sampling lower bounds with large error, close to 1, namely data-structure lower bounds. The large statistical distance corresponded to redundancy. In this paper we put forth another application to *communication protocols.* Here the large statistical distance corresponds to communication. We consider the following communication protocols: we associate to each output query a *party*. In addition to probing input cells as before, the parties also communicate. While sampling protocols have been studied before, see e.g. [ASTS+03, GW20], our setting where each party is charged to access input cells does not seem to have been studied before. Next we define it and then state our result. The result is an easy corollary and our main goal here is to give another interpretation of sampling lower bounds with large statistical distance.

Definition 11. A *sampler protocol* over Σ^m with word size w, q probes, and c total communication is a communication protocol among m parties. The parties share a public random string of cells of w bits each. At each point in time, the protocol specifies which Party i is to go next. Party i can either probe a cell, broadcast communication, or output a value in Σ and stop. The action of Party i at time t depends only on the values of the cells Party i probed in previous times, and on the communication transcript. The output of the protocol is the tuple of elements output by the parties. The number of probes made by each party is at most q, while c is a bound on total communication.

To get a sense of the parameters, consider for example $\mathrm{RANK}(U)$. We can sample it with no error with 1 probe and communication m/w (each party probes a different cell and broadcasts it – then the players sample exactly). And as we remarked earlier, it can also be sampled with $o(\log m)$ probes and no communication, up to constant error. We obtain the following lower bound, which interpolates between these two extremes.

Corollary 12. *Let Π be a sampler protocol with word size w, q probes, and communication c whose output has statistical distance δ from $\mathrm{RANK}(U)$. Then $c \geq m/w^{O(q)} + \log(1-\delta) - O(1)$.*

2 Techniques

Our results rely on a new proof technique which we call the *cell-probe sampling separator*, or just *separator* for brevity, and which is a main technical contribution of this work. Roughly speaking, this separator result says that if $f : W^{\mathbb{N}} \to \Sigma^m$ is a low-depth forest whose output distribution is close to a target distribution S over Σ^m, then we can restrict the input space to a subset $D \subseteq W^{\mathbb{N}}$ such that when the input to f comes from D, many trees in the output distribution $f(D)$ are *nearly pairwise independent*, and at the same time the output distribution is still not very far from the target S. This latter feature will be formalized by requiring that $f(D)$ is supported on a subset of the support of S, and has entropy almost equal to that of S.

A critical feature of the separator is that the number of trees that are guaranteed to be almost pairwise independent in $f(D)$ is much larger than the entropy gap between $f(D)$ and S. Formally, for a sufficiently spaced-out increasing sequence of integers $t_0, t_1, \ldots$, the separator will guarantee that for some value k there is a set $D_k = D$ and t_k trees that are nearly pairwise independent over $f(D_k)$, while the entropy gap is only about t_{k-1}. (The separator can also guarantee almost ℓ-wise independence for $\ell > 2$, but we only need $\ell = 2$ in our results.)

After some definitions we state the separator.

Definition 13. [Almost pairwise independence] Jointly distributed random variables X, Y are ϵ-independent if (X, Y) is ϵ-close in statistical distance to (X, Y') where Y' has the same distribution of Y and is independent from X.

The *min-entropy* $H_\infty(X)$ of a random variable X is $\min_a \log_2(1/\mathbb{P}[X = a])$.

Notation. To avoid clutter in the more technical exposition of the results, we adopt the convention that for a set S we also denote by S the uniform distribution U_S over S. The meaning will be clear from the context. For example, we shall simply write $\Delta(f(W^{\mathbb{N}}), S)$ for $\Delta(f(U_{W^{\mathbb{N}}}), U_S)$.

Theorem 14. *[Cell-probe sampling separator] There exists an integer $c \geq 1$ such that the following holds:*

Hypothesis: *Let $f : W^{\mathbb{N}} \to \Sigma^m$ be a depth-q forest with word size w. Let $\alpha \leq 1/c$. Let $t_0, t_1, \ldots$ be a sequence of integers with $t_i \geq t_{i-1} \cdot cqw/\alpha$ for every i. Let $S \subseteq \Sigma^m$ be a set and suppose that $\Delta(f(W^{\mathbb{N}}), S) \leq 1 - 2^{-t_0}$ where $2^{-t_0} \geq \sqrt{8/|S|}$.*

Conclusion: *There exists* $k, 1 \le k \le O(q/\alpha)$, $D_k \subseteq W^{\mathbb{N}}$, *and* t_k *indices* $T \subseteq [m]$ *such that:*
(0) $H_\infty(f(D_k)) \ge H_\infty(S) - t_{k-1} \cdot O(qw/\alpha)^2$;
(1) The support of $f(D_k)$ *is contained in* S*;*
(2) For every $i, j \in T$ *the random variables* $(f_i(D_k), f_j(D_k))$ *are* $O(\alpha)$*-independent.*

For example, we can set $t_i = m/w^{a(q/\alpha)-bi}$ which for suitable a, b and $q, w \le \log m$ satisfies the hypothesis.

Proof sketch of the separator. First we need to understand what it means for $\Delta(f(W^{\mathbb{N}}), S)$ to be at most $1-\epsilon$. One special case in which this happens is if the distribution $f(W^{\mathbb{N}})$ is equal to the uniform distribution over S with probability ϵ, and otherwise is say a fixed value. Our first Lemma 15 shows that this special case, more or less, is in fact the general case. Specifically, we can condition the input to f on an event of probability about ϵ so that, if D is the resulting set of inputs, $f(D)$ is supported inside of S, and the entropy of $f(D)$ is almost maximum.

At this point we forget S and our goal is to further restrict D so that we have many pairwise independent queries, and at the same time we do not lose too much in entropy.

First we apply the so-called *fixed-set lemma* from [GSV18]. This lemma shows that it is possible to moderately restrict D to a subset $D_1 \subseteq D$ so that no low-depth tree can distinguish D_1 from a *product distribution* R.

At this point, we ask if in $f(D_1)$ there are many (t_1) queries (a.k.a. trees) such that any two of them *intersect probes* with probability $\le \alpha$. Here we say that two trees intersect probes if there exists i such that both trees probe word i.

If the answer is positive: we argue that we are done. Let us explain why that is the case. First, we can write the probability that two queries probe the same word as a low–depth tree. By the fixed-set lemma, this probability is the same over D_1 and over the product distribution R. However, over a product distribution two queries are independent unless they probe the same word, hence over R two queries are α-independent, and it follows that the same is true over D_1.

We note that our use of the fixed-set lemma is different from [GSV18]. In the latter paper it was used to argue that the input to a tree looks uniform. By contrast, we use it to establish *pairwise independence* among trees, and critically we use it to bound the probability that two trees probe the same word.

If the answer is negative: In this case, by a version of simple "covering arguments" which are widespread since at least the sunflower lemma [ER60], there is a small set T of trees such that any other tree intersects probes with some tree in the set with probability $\ge \alpha$. Now the idea is to fix the probes of the trees in T to obtain a new input D_2 over which the total expected probe time is reduced. Then again we can apply the fixed-set lemma, and iterate the argument.

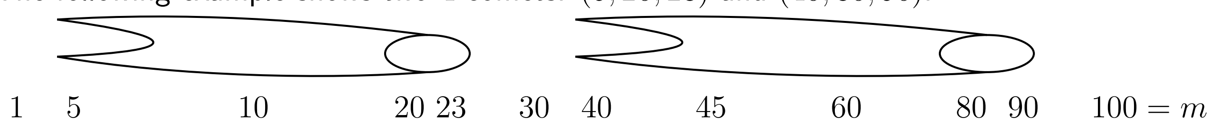
This fixing of the probes in T is inspired by a fixing that occurs in the data-structure lower bound for $\textsc{Rank}$ [PV10]. However, we note that our argument is different. The proof in [PV10] selects trees in a structured way, with a precise sequence of "gaps." By contrast, our selection comes from the covering argument and is, at this stage, unstructured: we simply count queries. More generally, the proof in [PV10] proceeds by an *encoding argument,* as is typical in data-structure lower bounds, which is tailored to the problem at hand. The separator avoids that and allows us to establish an intrinsic property of efficient samplers and data structures.

This concludes the informal overview. The formal proof is in Section 3.

Comparison with switching lemmas. *Switching lemmas* [FSS84, Ajt83, Yao85, Hås87, SBI04, Raz15, IMP12, Hås14] show that small-width DNF simplify under random restrictions. Since a depth-q decision tree over alphabet $\{0,1\}^w$ can be written as a DNF with width qw, switching lemmas apply to our model too. A main difference between switching lemmas and our separator is that the former restrict the input space *aggressively*, for example fixing all but a constant fraction of the input bits, while our separator restricts the input *moderately*, for example fixing a small, sub-linear number of input words. This distinction is critical, since our problems are easy for DNF.

Comets. Having established the separator, there remains to use it to prove lower bounds. Our approach is based on a combinatorial object that we call *comet*. A d-comet is a triple of integers where the first two, the comet's *tail,* are $\geq d$ times farther apart than the last two, the comet's *head.* We can imagine the sun at position ∞: Blown by solar winds, comet tails point away from the sun.

The following example shows two 4-comets: $(5, 20, 23)$ and $(40, 80, 90)$:



We show in Section 4 that any large set of integers contains many non-overlapping d-comets, for large enough d. In the proofs of the sampling lower bounds (Sections 6 and 7), this result is applied to the t_k trees given by the separator theorem. Because the entropy gap of D_k and S is, as remarked earlier, much less than t_k, it follows that we can find among the trees a comet that is "random," that is, roughly, the query outputs have a lot of entropy. However, we prove that this is impossible, because over $f(D_k)$ the queries are nearly independent, but we show that they are not so in (any restriction of) the target distribution. Here is where we use the geometry of comets: the long tail will impose correlations on the head of the comet. The way this is formalized depends on the problem. For $\text{CMP}\textsc{red}$, we can find blocks in Π which are just a little longer than comets' heads, guaranteeing correlations between the queries in the head. For $\textsc{Rank}$ the argument is a little more complex because a query depends on the entire prefix, so we shall need to guarantee that the bits corresponding to the comet have sufficiently high entropy even conditioned on the prefix.

2.1 Conclusion and open problems

This paper adds new tools to the study of sampling lower bounds, especially the separator theorem. Using them, a number of new lower bounds and separations are obtained. Several natural questions remain open. One is separating adaptive from non-adaptive samplers. Another is proving cell-probe lower bounds for sampling other distributions, such as *permutations, cf.* [Vio20]. The parameters of the separator do not seem strong enough for the latter goal; in brief, one would need to set α too small.

These new tools can also be used to generalize previous data-structure lower bounds, such as the one for $\textsc{Rank}$ [PV10], to *sampling* lower bounds. This additional information could be useful in understanding which techniques are suitable for further progress. For example, $\textsc{Membership}$ [Mil99, Tho13] is a long-standing problem in data structures which asks to store a subset of $[m]$ of size say $m/4$ so that membership queries can be computed fast. It is interesting to note that the corresponding sampling problem is easy: we can sample somewhat well the uniform distribution over these subsets in time $O(1)$ *using 2m input bits*. (Simply taking the And of adjacent pairs of bits will generate exactly the uniform distribution over m iid variables each coming up 1 with probability $1/4$; and this distribution has statistical distance only $1-\Omega(1/\sqrt{m})$ from the subsets.) Hence, unlike $\textsc{Rank}$, a strong lower bound for $\textsc{Membership}$ must exploit that the input length is bounded, and this might indicate why this problem is harder than $\textsc{Rank}$.

3 Proof of the separator Theorem 14

First we need to understand what it means to have slightly non-trivial statistical distance. Let P be a distribution over Σ^m. One way in which P can have statistical distance $\leq 1-\epsilon$ from S is if P is distributed like S with probability ϵ, and it is say fixed with probability $1-\epsilon$. In this case, P has actually very high entropy ($\log_2|S|$) conditioned on an event of probability ϵ. The next lemma shows that this in fact always happens.

Lemma 15. *Let P be a distribution over Σ^m and let $S \subseteq \Sigma^m$. Suppose that $\Delta(P,S) \leq 1-\epsilon$, where $\epsilon \geq \sqrt{8/|S|}$. Then there is a subset $S_0 \subseteq S$ of probability $\mathbb{P}_P[S_0] = \Omega(\epsilon)$ such that the distribution P conditioned on $P \in S_0$ has min-entropy $\geq H_\infty(S) - O(\log 1/\epsilon)$.*

Proof. We also write P for the random variable distributed according to P. Collect all the elements of S in increasing order of mass according to P until right before collecting cumulative mass $\epsilon/2$. Note we don't collect all of S, for else $\mathbb{P}[P \in S] \leq \epsilon/2$ and $\Delta(P,S) \geq 1-\epsilon/2$, contradicting the hypothesis.

Let β be the mass of the next element of S. Let S_0 be the collected elements, S_1 the rest of S, and T the complement of S. By definition, $\mathbb{P}[P \in S_0] < \epsilon/2$, and so $\mathbb{P}[P \in S_1 \bigcup T] \geq 1-\epsilon/2$. Also for every $x \in S_1$ we have $\mathbb{P}[P = x] \geq \beta$ and so $|S_1| \leq \beta^{-1}$. Combining these bounds with the assumption we have

$$1-\epsilon \geq \Delta(P,S) \geq \mathbb{P}[P \in S_1 \cup T] - \frac{|S_1|}{|S|} \geq 1-\epsilon/2 - \frac{\beta^{-1}}{|S|}$$

and so $\beta \leq 2/(\epsilon|S|)$.

Because we did not include in S_0 an element of mass β, and we only stop when we reach $\epsilon/2$, the mass of S_0 is $\geq \epsilon/2 - \beta \geq \epsilon/2 - 2/(\epsilon|S|)$. If $\epsilon \geq \sqrt{8/|S|}$ this mass is at least $\epsilon/4$.

For any $x \in S_0$ using the above bound on β we obtain

$$\mathbb{P}[P = x | P \in S_0] = \frac{\mathbb{P}[P = x]}{\mathbb{P}[P \in S_0]} \leq \frac{\beta}{\epsilon/4} \leq \frac{8}{\epsilon^2|S|},$$

as desired. ∎

The above lemma allows us to "forget" about S and focus on f. We need to show that we can restrict the input to a large subset such that many output trees are nearly independent. This is the content of the following theorem. To avoid having to think about infinite sets, in the remainder of the proof we set the input to the sampler to W^s for an integer s. This is without loss of generality, since obviously any forest of fixed depth can only access a finite number of input words.

We define the (entropy) *loss* of a subset $D' \subseteq D$ to be $\log_2(|D|/|D'|)$. So if D' contains half the elements of D the loss is one.

Theorem 16. *There exists an integer $c \geq 1$ such that the following holds:*

Hypothesis: *Let $f : W^s \to \Sigma^m$ be a depth-q forest with word size w. Let $\alpha \leq 1/c$. Let $t_0, t_1, \ldots$ be a sequence of integers with $t_i \geq t_{i-1} \cdot cqw/\alpha$ for every i. Let $D \subseteq W^s$ be a set with loss $\leq t_0$.*

Conclusion: *There exists $k, 1 \leq k \leq O(q/\alpha)$, $D_k \subseteq D$, and t_k indices $T \subseteq [m]$ such that:*
(1) The loss of $D_k \subseteq D$ is $\leq t_{k-1} \cdot (qw/\alpha)^2$;
(2) For every $i, j \in T$ the random variables $f_i(D_k), f_j(D_k)$ are $O(\alpha)$-independent.

Let us first show how this gives the separator Theorem 14.

Proof. [Proof of Theorem 14 from Theorem 16.] We apply Lemma 15 to $P = f(U)$. Given $S_0 \subseteq S$ from the lemma, we let $D \subseteq W^s$ be the preimage of S_0 according to f. By the lemma, $|D|/|W|^s \geq \Omega(2^{-t_0})$, that is, the loss of $D \subseteq W^s$ is $t_0 + O(1)$. Moreover, $H_\infty(f(D)) \geq \log|S| - O(t_0)$.

We now apply Theorem 16 to this set D and the sequence $t_0 + O(1), t_1, t_2 \ldots$. We can adjust the constant c so that this satisfies the hypothesis. The theorem gives $D_k \subseteq D$ with loss $\leq t_{k-1} \cdot (qw/\alpha)^2$.

Observe that the support of $f(D_k)$ is contained in S, because the support of $f(D)$ is $S_0 \subseteq S$ and $D_k \subseteq D$.

To verify the bound on $H_\infty(f(D_k))$, note that

$$\mathbb{P}[f(D) = x] \geq \mathbb{P}[f(D_k) = x]|D_k|/|D|.$$

Taking inverses and then logs we obtain

$$\log(1/\mathbb{P}[f(D) = x]) \leq \log(1/\mathbb{P}[f(D_k) = x]) + \log(|D|/|D_k|).$$

The left-hand side is at least $H_\infty(f(D)) \geq \log|S| - O(t_0)$. While $\log(|D|/|D_k|) \leq t_{k-1} \cdot (qw/\alpha)^2$. Hence,

$$\log(1/\mathbb{P}[f(D_k) = x]) \geq \log|S| - O(t_0) - t_{k-1} \cdot (qw/\alpha)^2,$$

for any x. The result follows. □

3.1 Proof of Theorem 16

The main technical lemma is the following one, which is like Theorem 16 but the requirement of independence is replaced by others easier to work with.

Lemma 17. *Theorem 16 holds if we replace (2) with:*

(2') for every $i, j \in T$: the probability over D_k that $f_i(D_k)$ and $f_j(D_k)$ don't make all distinct probes is $\leq \alpha$, and

(2") there exists a product distribution R over words (that is, the words are independent) such that for every depth-$2q$ tree g, $g(D_k)$ and $g(R)$ are α-close.

Lemma 18. *(2') and (2") in Lemma 17 imply (2) in Theorem 16.*

Proof. Let $X = D_k$. Think of $(f_i(X), f_j(X))$ as the output of the tree g obtained by appending f_j to the leaves of f_i. Note that g makes $2q$ probes, possibly repeated. By (2"), there is a product distribution R such that $g(X)$ and $g(R)$ are α-close. Also, the probability that g repeats a probe over X is α-close to the probability that it repeats it over R. Here we use that this probability can be written as the probability that a tree d of depth $2q$ outputs 1, and that the output distributions of d over X and R are α-close. (Tree d is obtained from g by replacing any repeated probe along any path with a leaf 1, and any other leaf with 0.)

By this and (2') the probability that g repeats a probe over R is $\leq 2\alpha$. Because R is product, as long as probes are not repeated the output distribution does not change if we answer the first q probes with R^1 and the next q probes with R^2 where R^1, R^2 are iid copies of R. This shows that $(f_i(X), f_j(X))$ is $O(\alpha)$-close to $(f_i(R^1), f_j(R^2))$. Using again (2"), we can replace each R^i with X^i, where X^1, X^2 are iid copies of X. This gives that $(f_i(X), f_j(X))$ is $O(\alpha)$-close to $(f_i(X^1), f_j(X^2))$. Adjusting constants concludes the proof. $\square$

3.2 Proof of Lemma 17

We shall be concerned with inputs in various subsets $X \subseteq D$. If an input word is constant for every $x \in X$ then it needs not be probed but can be "hardwired" in the trees. We shall assume that the trees are always simplified accordingly. We denote by $G(x, X)$ the total number of probes made by all trees on input x, where the trees are simplified with respect to X.

We use the following fixed-set lemma from [GSV18]. We say that distributions X and Y are α-indistinguishable by depth-q trees if for any such tree t, the statistical distance between $t(X)$ and $t(Y)$ is $\leq \alpha$.

Lemma 19. *[[GSV18], Lemma 3.14.] Let $B \subseteq W^s$ be a subset with loss $\leq b$, where $W = \{0,1\}^w$. There exists $B_1 \subseteq B$ and a product distribution R such that B_1 and R are α-indistinguishable by depth-$2q$ decision trees. Moreover, the loss of $B_1 \subseteq W^s$ is $\leq b \cdot O(wq/\alpha)$.*

For completeness we include the proof in Appendix A.

We begin by applying this lemma to D obtaining $D_1 \subseteq D$ with loss $t_0 \cdot O(wq/\alpha)$. This is the beginning of Iteration 1.

Our goal is to show that at the beginning of Iteration k there exists a subset $D_k \subseteq D$ enjoying the following properties:

(1) [in Theorem 16] the loss of D_k is $\leq t_{k-1} \cdot (qw/\alpha)^2$,

(2") [in Lemma 17] there exists a product distribution R over words such that for every depth-$2q$ tree g, $g(D_k)$ and $g(R)$ are α-close.

(3) $\max_{x \in D_k} G(x, D_k) \leq m(q - \alpha(k-1)/4)$.

Note all these hold at the beginning of Iteration 1.

In an iteration, collect as many trees as possible such that for any two of them, the probability over D_k that they intersect probes is $\leq \alpha$. If you have t_k, then (2') in Lemma 17 holds as well, concluding the proof.

Otherwise, you have a collection of t_k trees such that any other tree will intersect a probe with one of those t with probability $\geq \alpha$. We are going to use this to proceed to the next iteration, i.e., increase the value of k by 1. Because G is non-negative, Property (3) above implies that there can be at most $O(q/\alpha)$ iterations, as desired.

Write Y for the $\le t_k q$ words probed by the t_k trees in D_k. This is done according to a canonical order, and is a valid definition because the first probe of a tree is fixed, the second is fixed once the answer to the first is, and so on.

Support size. Let $D_{k,y}$ be the inputs in D_k with $Y = y$. We have

$$\mathbb{E}_Y[|D_k|/|D_{k,Y}|] = \sum_y \mathbb{P}[Y = y]\frac{|D_k|}{|D_{k,y}|} = \sum_y 1 \le |W|^{t_k q}.$$

By Markov's inequality $\mathbb{P}_Y[|D_k|/|D_{k,Y}| \ge M] \le |W|^{t_k q}/M$. And so with probability $\ge 1 - |W|^{t_k q}/M$ over Y we have $|D_{k,Y}| \ge |D_k|/M$.

Intersection. For a tree f_i let $I_i(x)$ equal 1 if on input x tree f_i intersects probes with at least one of the t_k trees collected, and equal 0 otherwise. Note that for every input x and fixing y we have

$$G(x, D_{k,y}) \le G(x, D_k) - \sum_{i\in[m]} I_i(x).$$

Because $\mathbb{P}_{x\in D_k}[I_i(x) = 1] \ge \alpha$ for every i, we have

$$\mathbb{E}_{x\in D_k}\left[\sum_{i\in[m]} I_i(x)\right] \ge \alpha m.$$

Because the inner sum is $\le m$, by Markov's inequality we have that with probability $\ge \alpha/2$ over the choice of Y

$$\mathbb{E}_{x\in D_{k.Y}}\left[\sum_{i\in[m]} I_i(x)\right] \ge \alpha m/2,$$

and

$$\mathbb{E}_{x\in D_{k.Y}} G(x, D_{k,Y}) \le \mathbb{E}_{x\in D_{k.Y}} G(x, D_k) - \alpha m/2.$$

Combining the arguments. Selecting $M = 2|W|^{t_k q}/\alpha$ above, and by a union bound, there is a value $\bar{y}$ so that

$$\mathbb{E}_{x\in D_{k,\bar{y}}}[G(x, D_{k,\bar{y}})] \le \mathbb{E}_{x\in D_{k,\bar{y}}} G(x, D_k) - \alpha m/2;$$

and at the same time $|D_{k,\bar{y}}| \ge |D_k| \cdot \alpha|W|^{-t_k q}/2$. That is, we increase the loss by

$$\le \log(1/\alpha) + w t_k q + 1.$$

Recall that the loss of D_k is $t_{k-1} \cdot (qw/\alpha)^2$.

Note that $D_{k,\bar{y}}$ is still uniform over its support, since it is D_k conditioned on a particular choice for $\le t_k q$ words. Even though the words are chosen adaptively in D_k, once we condition on a particular value, their locations are fixed.

Reducing G for every input. By Markov's inequality,

$$\mathbb{P}_{x\in D_{k,\bar{y}}}[G(x,D_{k,\bar{y}})\geq(\mathbb{E}_{x\in D_{k,\bar{y}}}G(x,D_k)-\alpha m/2)(1+\alpha/(4q))]\leq\frac{1}{1+\alpha/(4q)}\leq 1-\alpha/(8q).$$

Hence, for $\geq\alpha/(8q)$ fraction of the inputs x in $D_{k,\bar{y}}$ we have

$$\begin{aligned}G(x,D_{k,\bar{y}})&\leq(\mathbb{E}_{x\in D_{k,\bar{y}}}G(x,D_k)-\alpha m/2)(1+\alpha/(4q))\\&\leq m(q-\alpha(k-1)/4)(1+\alpha/(4q))-\alpha m/2\\&\leq m(q-\alpha k/4),\end{aligned}$$

using (3) in the second inequality. Let $D'_{k,\bar{y}}$ be the set of these inputs. The above gives the desired bound on $\max_{x\in D'_{k,\bar{y}}}G(x,D'_{k,\bar{y}})$, and note that the loss of $D'_{k,\bar{y}}\subseteq D_{k,\bar{y}}$ is $\leq\log(8q/\alpha)$.

Fixed-set lemma. Finally, we apply the fixed-set lemma to $D'_{k,\bar{y}}$ to obtain D_{k+1}; this gives (2"). This application multiplies the loss by $O(wq/\alpha)$, bringing the loss of $D_{k+1}\subseteq D$ to

$$O(wq/\alpha)\cdot O(t_{k-1}\cdot(qw/\alpha)^2+\log(1/\alpha)+wt_kq+1).$$

We need this loss to be at most $t_k\cdot(qw/\alpha)^2$. Dividing by wq/α we need to verify that

$$O(t_{k-1}\cdot(qw/\alpha)^2)+O(\log 1/\alpha)+O(wt_kq)+O(1)\leq t_k\cdot qw/\alpha.$$

We claim that each term on the left-hand side is at most one-fourth of the right-hand side. For the first term we use the hypothesis that $t_k\geq t_{k-1}\cdot cqw/\alpha$ for a large enough c, and for the third we use that $\alpha\leq 1/c$ and pick c large enough. This gives (1).

Because $D_{k+1}\subseteq D'_{k,\bar{y}}$, the bound on G still holds for D_{k+1}, and this gives (3).

4 Comets

In this section we define comets and prove a comet-finding lemma which will be used in our sampling lower bound.

Definition 20. *A d-comet is a triple of indices (i,j,k) from $[m]$ with $i<j<k$ such that $j-i\geq d(k-j)$. We call (j,k) the* head *and (i,j) the tail. The head length is $k-j$. A set of comets $\{(i_h,j_h,k_h)\}_h$ is* disjoint *if the intervals $[i_h,k_h]$ are disjoint.*

Lemma 21. *[Comet-finding] A subset of $\{1,2,\ldots,m\}$ of size m/ℓ^b contains $\geq m/\ell^{b+c+O(1)}$ disjoint ℓ^c-comets where the head lengths are all in $[\ell^h,\ell^{h+1}]$ for some integer $h\leq b+c+O(1)$, for any $m,b\leq\ell,c\leq\ell$, and $\ell\geq\log m$.*

Proof. Let $d=\ell^c$. First we claim that any subset of size $n:=d\log m+2$ contains a d-comet. Let the elements in the set be $a_1,a_2,\ldots$ in increasing order. If (a_1,a_2,a_3) is not a d-comet then $a_3-a_2>(a_2-a_1)/d$, and so $a_3-a_1=a_3-a_2+a_2-a_1\geq(a_2-a_1)(1+1/d)$. Then again if (a_1,a_3,a_4) is not a d-comet we have $a_4-a_3\geq(a_3-a_1)/d$ and so $a_4-a_1\geq(a_3-a_1)(1+1/d)\geq$

$(a_2 - a_1)(1 + 1/d)^2$. If we continue this way $n-2$ times, we obtain $a_n \geq (1 + 1/d)^{n-2} > m$, which is a contradiction.

Now divide the $t := m/\ell^b$ elements of the given set into consecutive blocks of size n. By the previous paragraph, each block contains a comet. Hence we have $\geq t/n - 1$ disjoint d-comets.

At least half of these comets have heads of length $\leq O(mn/t) = \ell^{b+c+O(1)}$, otherwise half the comets have heads longer than that, and we run out of space. Let C_i be the subset of these comets whose head length is in $[\ell^i, \ell^{i+1})$. We only need to consider $i \leq b + c + O(1)$. Hence, there exists $i = h$ and

$$\Omega\left(\frac{t}{n}\right)\frac{1}{b+c+O(1)} \geq \frac{m}{\ell^{b+c+O(1)}}$$

disjoint comets with head lengths in $[\ell^h, \ell^{h+1})$, using that both b and c are $\leq \ell$. □

5 A lemma about entropy

In this section we quickly recall a basic result about entropy which will be used in our sampling lower bounds. The *entropy* H of a random variable X is defined as $H(X) := \sum_x \Pr[X = x] \cdot \lg(1/\Pr[X = x])$. The conditional entropy $H(X|Y) := E_{y \in Y} H(X|Y = y)$ (cf. Chapter 2 in [CT06]).

Lemma 22. *Let $Z = (Z_1, \ldots, Z_k)$ where Z_i is supported over a set S_i, and let $\sum_i \log|S_i| = M$. Suppose $H(Z) \geq M - a$. There is a set $G \subseteq [k]$ of size $|G| \geq k - a/\epsilon$ such that for any $i \in G$ we have*

$$H(Z_i|Z_1 Z_2 \ldots Z_{i-1}) \geq \log|S_i| - \epsilon.$$

In particular, Z_i is $4\sqrt{\epsilon}$ close to uniform over S_i.

Proof. By the chain rule for entropy ([CT06], Equation 2.21)

$$\sum_{i \leq k} (\log|S_i| - H(Z_i|Z_1 Z_2 \ldots Z_{i-1})) \leq a.$$

Applying Markov inequality to the non-negative random variable $\log|S_i| - H(Z_i|Z_1 Z_2 \ldots Z_{i-1})$ (for random $i \in [k]$), we have

$$\mathbb{P}_{i \in [k]}[\log|S_i| - H(Z_i|Z_1 Z_2 \ldots Z_{i-1}) \geq \epsilon] \leq a/(k \cdot \epsilon),$$

yielding the desired G.

The "in particular" part holds because conditioning reduces entropy: $H(Z_i) \geq H(Z_i|Z_1 Z_2 \ldots Z_{i-1})$ *(*[CT06], Equations 2.60 and 2.92) and then applying Pinsker's inequality ([CK82], Chapter 3; Exercise 17). □

6 Proof of Theorem 8

We can assume that $q \leq w$, for else the statistical bound is trivial and the theorem is true. We apply Theorem 14 with $\alpha = 1/10$ and the sequence

$$t_i := m/w^{c_0(q/\alpha) - c_1 i},$$

for constants c_0, c_1 to be set later. For large enough c_1 this satisfies the hypothesis of the theorem that $t_i \geq t_{i-1} \cdot cqw/\alpha$. We also need to show that $2^{-t_0} \geq \sqrt{8/|S|}$, where $|S|$ is the number of matrices Π in the definition of $\mathrm{CMP}\textsc{red}$. This is true since $|S| \geq 2^{\Omega(m/w)}$.

Let k, D_k, and t_k be as provided by the theorem. Recall that

$$H_\infty(f(D_k)) \geq H(\Pi) - t_{k-1} \cdot O(qw/\alpha)^2.$$

Finding comets among trees. The theorem provides $t := t_k = m/w^{c_0(q/\alpha)-c_1 k}$ trees. Applying the Comet-Finding Lemma 21 with $c = 3$ and $\ell = w \geq \log m$ gives a set of

$$t' := m/w^{c_0(q/\alpha)-c_1 k+O(1)}$$

disjoint w^3-comets, where the head lengths are in $[w^h, w^{h+1})$ for some $h \leq c_0(q/\alpha)+O(1)$. Note that to apply the lemma we need that $c_0(q/\alpha) \leq w$. This is guaranteed since $w \geq \log m$ and $q = O(\log m)/\log\log m$ for else the conclusion of the theorem holds trivially.

We shall get a contradiction looking at the row of the matrix corresponding to blocks of length w^{h+2}; the other rows can be ignored.

A random comet. To each of the above t' comets we associate three *relevant*, consecutive blocks. Of these, the middle block is the first block that intersects the head of the comet. Note that:

– the relevant blocks cover the head of the comet, since the blocks have length w^{h+2} while the head has length $\leq w^{h+1}$.

– the relevant blocks of different comets are disjoint, since the tails of each comet have length $\geq w^h \cdot w^3$, while the blocks relevant to a comet are contained in an interval of length $3w^{h+2}$ intersecting the head.

Note that from $\mathrm{CMP}\textsc{red}(\Pi)$ we can reconstruct Π, and moreover $f(D_k)$ is in the range of $\mathrm{CMP}\textsc{red}$. Hence we can define

$$X := \mathrm{CMP}\textsc{red}^{-1}(f(D_k))$$

and we have $H(X) = H(f(D_k))$. Let B_i be the portion of X in the three blocks relevant to comet i, in our current set of t' comets. Recall that in row $h+2$ of the $\mathrm{CMP}\textsc{red}$ distribution Π, each block is given by a variable uniform over a support of size $2 \cdot w^{h+2}$. Hence B_i is a random variable uniform over its support $\mathrm{Supp}(B_i)$ of size $(2 \cdot w^{h+2})^3$.

We want to argue that one such variable is close to uniform in our distribution $f(D_k)$. Indeed, recall from the beginning of the proof that

$$H(X) \geq H_\infty(f(D_k)) \geq H(\Pi) - t_{k-1} \cdot O(qw/\alpha)^2.$$

Since $H(X,Y) \leq H(X) + H(Y)$ for any random variables X, Y, we have that,

$$H(B_1, B_2, \ldots, B_{t'}) \geq t' \log|\mathrm{Supp}(B_i)| - t_{k-1} \cdot O(qw/\alpha)^2.$$

By Lemma 22, each B_i is α-close to uniform, except for those in a "forbidden" set of size $t_{k-1} \cdot O(q^2w^2/\alpha^4)$.

Now for the critical point, t' is larger than the size of this forbidden set. This is true because we only lost $w^{O(1)}$ factors, so it suffices to make the constant c_1 large enough in the definition of the sequence t_i. Formally,

$$t_{k-1} \cdot O(q^2 w^2/\alpha^4) = (m/w^{c_0(q/\alpha)-c_1(k-1)}) \cdot O(q^2 w^2/\alpha^4)$$

which is smaller than $t' = m/w^{c_0(q/\alpha)-c_1 k+O(1)}$ for c_1 large enough. Here we are using that $q \leq O(\log m)/\log\log m, w \geq \log m, \alpha = \Theta(1)$.

Breaking correlation in the random comet. At this point we have a w^3-comet (p, i, j) where the head length $(j - i)$ is in $[w^h, w^{h+1}]$ and

(1) The answers to queries i and j are α-independent, and

(2) the relevant blocks are α-close to uniform.

In the query answers consider just the color corresponding to row $h+2$ for query i and j. Let them be $C(i)$ and $C(j)$.

Because the relevant blocks are α-close to uniform, for any color c we have both $\mathbb{P}[C(i) = c] \leq 1/2 + \alpha$ and $\mathbb{P}[C(j) = c] \leq 1/2 + \alpha$. Also, because $C(i)$ and $C(j)$ are α-independent, we have $\mathbb{P}[C(i) = C(j)] \leq 1/2 + 2\alpha$.

However, $C(i)$ and $C(j)$ are in fact highly correlated. The only event in which $\mathbb{P}[C(i) \neq C(j)]$ is if the head of the comet contains an element. The head has length $\leq w^{h+1}$. The blocks have length w^{h+2}. If the variables in the blocks were uniform, the chance that the head contains an element is $\leq 1/w$. The block is only α-close to uniform, so this probability is $\leq 1/w + \alpha$. Hence, $\mathbb{P}[C(i) = C(j)] \geq 1 - 1/w - \alpha$. For $\alpha = 1/10$, this is larger than the above value of $1/2 - 2\alpha$, concluding the proof.

Reducing CMPred to Pred. We quickly recall this reduction to justify the claim made in the introduction that we obtain a lower bound for $\textsc{Pred}$ under a suitable distribution. Given $x, y \in \{0,1\}^m$ we create $z \in \{0,1\}^{m^3}$ such that $(\textsc{Pred}(x)_i, \textsc{Pred}(y)_i)$ depends only on (and therefore can be reduced to computing) $\textsc{Pred}(z)_j$. Let $x \otimes y$ be the $m \times m$ matrix where the i, j coordinate is $x_i \cdot y_j$. We can also think of this as a vector z in $\{0,1\}^{m^2}$ listing the elements in the matrix in row order. Note that $(\textsc{Pred}(x)_m, \textsc{Pred}(y)_m)$ is the same as $\textsc{Pred}(z)_{m^2}$ written in base m. However to compute $(\textsc{Pred}(x)_i, \textsc{Pred}(y)_i)$ for $i < m$ this doesn't quite work. One simple fix is to *zero-out* part of the matrix. Define $x \otimes_i y$ to be the same as $x \otimes y$ except that only the top-left $i \times i$ sub-matrix may be non-zero; Then $(\textsc{Pred}(x)_i, \textsc{Pred}(y)_i)$ can be obtained from $\textsc{Pred}(x \otimes_i y)_{i \cdot m^2}$. Hence we can reduce two instances x and y of $\textsc{Pred}$ to the instance $(x \otimes_1 y, x \otimes_2 y, \ldots)$. Repeat ℓ times for 2^ℓ instances.

7 Proof of Theorem 4

We can assume that $q \leq \log m$, for else the statistical statistical bound is trivial and the theorem is true. We apply the separator Theorem 14 with $\alpha = 1/1000$ and the sequence

$$t_i := m/w^{c_0(q/\alpha)-c_1 i},$$

for constants c_0, c_1 to be set later. For large enough c_1 this satisfies the hypothesis of the theorem that $t_i \geq t_{i-1} \cdot cqw/\alpha$. The hypothesis that $1 - 2^{-t_0} \geq \sqrt{8/|S|} = \sqrt{8/2^m}$ holds as well since $|S| = 2^m$.

Let k and D_k be as given by the theorem. Let

$$X := \text{RANK}^{-1} f(D_k).$$

Note that this is a valid definition because $f(D_k)$ is in the range of RANK, and the latter is 1-1. The separator theorem guarantees that $H_\infty(f(D_k)) \geq m - t'_{k-1}$, where

$$t'_{k-1} := t_{k-1} \cdot O(q^2 w^2/\alpha^4).$$

Hence also $H(X) \geq m - t'_{k-1}$.

Comets. We now apply the comet-finding Lemma 21 to the $t_k = m/w^{c_0(q/\alpha) - c_1 k}$ trees given by the separator. For $c = 1$, the lemma gives a set of

$$t'_k := m/w^{c_0(q/\alpha) - c_1 k + O(1)}$$

disjoint w-comets. We shall only use that they are 100-comets, and their head lengths will not be relevant now. We want to find a comet whose outputs are "sufficiently random."

Define $a := t'_{k-1}$ and $b := t'_k$.

Partition X into b consecutive blocks, where each block contains exactly one comet and intersects no others. Let $Z_1, Z_2, \ldots, Z_b$ be the blocks, and let $|Z_i| = s_i$ with $\sum_i s_i = m$. Applying Lemma 22 we find $\geq b - a/\epsilon$ blocks i such that $H(Z_i | Z_1 Z_2 \ldots Z_{i-1}) \geq s_i - \epsilon$. We set $\epsilon = 1/w$ (a sufficiently small constant would be enough), and we verify that $b - a/\epsilon \geq 1$, yielding at least one block i^* such that

$$H(Z_{i^*} | Z_1 Z_2 \ldots Z_{i^*-1}) \geq s_{i^*} - \epsilon. \tag{1}$$

The inequality $b - a/\epsilon \geq 1$ is true because we only lost $w^{O(1)}$ factors, so it suffices to make the constant c_1 large enough in the definition of the sequence t_i. Formally,

$$\frac{b}{a} = \frac{t_k}{t_{k-1}} \cdot \frac{1}{O(q^2 w^2/\alpha^4) \cdot w^{O(1)}} \geq \frac{w^{c_1}}{w^{O(1)}} > w^{100}.$$

The inequalities holds for c_1 large enough and using $q \leq \log m, w \geq \log m, \alpha = \Theta(1)$.

Breaking correlation in the random comet. Hence we now have a comet (p, i, j) that is contained in an interval Z_{i^*} such that:

(1) Equation 1 holds, and

(2) $f_i(D_k), f_j(D_k)$ are α-independent.

The next lemma directly contradicts this and concludes the proof.

Lemma 23. *Let $X_1, X_2, \ldots, X_m$ be $0-1$ random variables, and (p, i, j) a c-comet for a sufficiently large c. Let $\ell := i - p$ and $d := j - i$.*

Suppose that

$$H(X_{p+1}, X_{p+2}, \dots, X_j | X_1, X_2, \dots, X_p) \geq \ell + d - 1/c.$$

Then there exists an integer t such that

$$\mathbb{P}_X\left[\textsc{Rank}(X)_j \geq t + \ell/2 + d/2 + c^{1/3}\sqrt{d}\right] \geq 1/10, \text{ and}$$
$$\mathbb{P}_X\left[\textsc{Rank}(X)_i < t + \ell/2\right] \geq 1/10, \text{ but}$$
$$\mathbb{P}_X\left[\textsc{Rank}(X)_j \geq t + \ell/2 + d/2 + c^{1/3}\sqrt{d} \bigwedge \textsc{Rank}(X)_i < t + \ell/2\right] \leq 1/1000 (\ll 1/10 \cdot 1/10).$$

Proof. Let us start with the last inequality, because we can prove it without getting our hands on t. The probability is at most

$$\mathbb{P}_X\left[\sum_{k=i+1}^{j} X_k \geq d/2 + c^{1/3}\sqrt{d}\right].$$

By Pinsker's inequality ([CK82], Chapter 3; Exercise 17) the distribution of $X_{i+1}, X_{i+2}, \dots, X_j$ is $4/\sqrt{c}$ close to the uniform $U_1 U_2 \dots U_d$. Hence the above probability is

$$\leq \Pr_U\left[\sum_{k=1}^{d} U_k \geq d/2 + c^{1/3}\sqrt{d}\right] + 4/\sqrt{c} \leq 1/2000 + 4/\sqrt{c} \leq 1/1000.$$

where the second inequality follows from Chebyshev's inequality for sufficiently large c.

We now verify the first two inequalities in the conclusion of the lemma. Let $Y := X_1, X_2, \dots, X_p$ stand for the prefix, and $Z := X_{p+1}, X_{p+2}, \dots, X_j$ for the $\ell + d$ high-entropy variables. Let

$$A := \{y \in \{0,1\}^p : H(Z|Y = y) \geq \ell + d - 2/c\}$$

be the set of prefix values conditioned on which Z has high entropy. We claim that $\mathbb{P}[Y \in A] \geq 1/2$. This is because, applying Markov Inequality to the non-negative random variable $\ell + d - H(Z|Y = y)$ (for y chosen according to Y),

$$\begin{aligned}\mathbb{P}[Y \notin A] = &\mathbb{P}_{y \in Y}[\ell + d - H(Z|Y = y) > 2/c] \\ \leq &\mathbb{E}_{y \in Y}[\ell + d - H(Z|Y = y)]/(2/c) \\ = &(\ell + d - H(Z|Y))/(2/c) \leq (1/c)/(2/c) = 1/2.\end{aligned}$$

Note that for every $y \in A$ we have, by definition, that the $(\ell + d)$-bit random variable $(Z|Y = y)$ has entropy at least $\ell + d - 2/c$, and so by Pinsker's inequality ([CK82], Chapter 3; Exercise 17) the random variable $(Z|Y = y)$ is $(\epsilon := 4\sqrt{2/c})$-close to uniform over $\{0,1\}^{\ell+d}$. Therefore, for any subset $S \subseteq A$, the random variable

$$(Z|Y \in S) \text{ is } \epsilon\text{-close to uniform over } \{0,1\}^{\ell+d}. \tag{2}$$

Now define t to be the largest integer such that

$$\mathbb{P}\left[Y \in A \wedge \textsc{Rank}(X)_p \geq t\right] \geq 1/4. \tag{3}$$

Since by definition of t we have $\mathbb{P}[Y \in A \wedge \textsc{Rank}(X)_p \geq t+1] < 1/4$, we also have

$$\mathbb{P}\left[Y \in A \wedge \textsc{Rank}(X)_p \leq t\right] \geq 1/2 - 1/4 = 1/4. \tag{4}$$

We obtain the desired conclusions as follows, denoting by $U_1, U_2, \ldots,$ uniform and independent $0-1$ random variables. The first probability in the conclusion of the lemma is at least

$$\mathbb{P}\left[\textsc{Rank}(X)_j \geq t + (\ell+d)/2 + \sqrt{\ell}/c^{1/6}\right]$$

because $\ell \geq c \dot{\cdot} d$.

Writing $\textsc{Rank}(X)_j$ as the sum of the first p bits and the rest, the above probability is at least

$$\mathbb{P}[\sum_{k\leq \ell+d} Z_k \geq (\ell+d)/2 + \sqrt{\ell}/c^{1/6} | Y \in A \wedge \textsc{Rank}(X)_p \geq t] \cdot \mathbb{P}[Y \in A \wedge \textsc{Rank}(X)_p \geq t].$$

The second factor is $\geq 1/4$ by (3). Also by (2) in the first factor we can replace the Z_k with uniform bits changing the probability by at most ϵ. Hence the first factor is at least

$$\mathbb{P}[\sum_{k\leq \ell+d} U_k \geq (\ell+d)/2 + \sqrt{\ell}/c^{1/6}] - \epsilon.$$

In turn the probability is

$$\geq 1/2 - \sqrt{\ell}/c^{1/6} \cdot \Theta(1/\sqrt{\ell}) \geq 1/2 - \Theta(1/c^{1/6})$$

using an estimate of the central binomial coefficient provided e.g. in [CT06], Lemma 17.5.1.

Overall, the first probability in the conclusion of the lemma is

$$\left(1/2 - \Theta(1/c^{1/6}) - \epsilon\right)(1/4) \geq 1/10$$

for large enough c.

We now turn to the second probability in the conclusion of the lemma. Proceeding in a similar way, this probability is at least

$$\begin{aligned}&\mathbb{P}\left[\sum_{k\leq \ell} Z_k < \ell/2 \middle| Y \in A \wedge \sum_k Y_k \leq t\right] \cdot \mathbb{P}\left[Y \in A \wedge \sum_k Y_k \leq t\right]\\ &\geq \left(\mathbb{P}\left[\sum_{k\leq \ell} U_k < \ell/2\right] - \epsilon\right) \cdot (1/4) \geq (1/2-\epsilon)\cdot(1/4) \geq 1/10\end{aligned}$$

for all sufficiently large c. Here the second inequality uses (2) and (4). $\square$

8 Proof of Corollary 12

We claim that there is a depth-q $f : W^s \to \Sigma^m$ such that the statistical distance between $f(U)$ and S is at most $1 - \Omega(1-\delta)/2^c$. Then the result follows from the sampling lower bounds.

To prove the claim, consider fixing the communication transcript of the protocol to i. Because the communication is fixed, each party can be implemented as a depth-q forest. If the protocol dictates Party j to send a message that does not match i, Party j outputs any value and stops. Let $C = 2^c$ and consider the C forests f_i where each f_i corresponds to the protocols run with fixed communication transcript i. The result now follows from the next lemma, letting P_i be the output distribution of f_i, and $Q_i(x)$ the probability that f outputs x using transcript i.

Lemma 24. *Let P be a distribution and S a set. Suppose $P(x) = \sum_{i=1}^{C} Q_i(x)$ where each $Q_i(x) \in [0,1]$ but $\sum_x Q_i(x) = 1$ is not required. Suppose $\Delta(P,S) = \delta$. Define P_i to be the probability distribution with $P_i(x) = Q_i(x)$ and the remainder $1 - \sum_x Q_i(x)$ mass is put arbitrarily.*

Then there exists i such that $\Delta(P_i, S) \le 1 - \epsilon$ where $\epsilon := 0.1 \cdot (1-\delta)/C$.

Proof. We use that
$$\Delta(A,B) = \sum_{x:A(x)\ge B(x)} A(x) - B(x).$$

Suppose there exists i such that
$$\sum_{x:Q_i(x)\le S(x)} Q_i(x) \ge \epsilon.$$

Then $\Delta(P_i,S) = \sum_{x:P_i(x)\le S(x)} S(x) - P_i(x) \le 1-\epsilon$ and we are done.
Let
$$T_i := \{x \in S : Q_i(x) \ge 1/|S|\} \subseteq S.$$

By above each Q_i puts mass at most ϵ outside of T_i. Now we show that the mass in T_i is concentrated on few points. Suppose $|T_i| \ge \epsilon|S|$ for some i. Then $\Delta(P_i,S) = \sum_{x:S(x)\le P_i(x)} P_i(x) - S(x) \le 1 - \sum_{x\in T_i} 1/|S| \le 1-\epsilon$ and we are done.
Now we can contradict the hypothesis. Let
$$T := \{x \in S : P(x) \ge 1/|S|\} \subseteq S.$$

Note that $T_i \subseteq T$ for every i. Hence each Q_i contributes $\le \epsilon|S|$ elements to T via T_i, and further contributes ϵ mass to distribute for others. With mass α we obtain $\le \alpha|S|$ elements such that $P(x) \ge S(x)$. Hence $|T| \le C\epsilon|S| + C\epsilon|S| = 2C\epsilon|S|$.

Let α, β, γ be respectively the masses that P puts outside of S, in T, and in $S \setminus T$. Note that $\gamma \le C\epsilon$, since each Q_i puts $\le \epsilon$ mass on $S \setminus T_i$. We have
$$\Delta(P,S) = \sum_{x:P(x)\ge S(x)} P(x) - S(x) = \alpha + \beta - \sum_{x\in T} S(x) \ge \alpha + \beta - 2C\epsilon.$$

We have $\alpha + \beta = 1 - \gamma \ge 1 - C\epsilon$.
Hence we get
$$\delta \ge \Delta(P,S) \ge 1 - 3C\epsilon,$$
as desired. □

9 A negative result for sampling slices

Let S be the set of m-bit strings with Hamming weight $m/3$, assuming 3 divides m. Recall we also denote by S the uniform distribution over the same set. In this section we prove the following negative result for sampling S.

Theorem 25. *Let $f : \{0,1\}^{\mathbb{N}} \to \{0,1\}^m$ be a depth-q forest with word size $w = 1$. Then $\Delta(f(U_{\{0,1\}^{\mathbb{N}}}),S) \ge 1 - 2 \cdot 2^{-\sqrt{m}/2^{O(q)}}$.*

The $\sqrt{m}$ term can be optimized but we do not do so for simplicity. This result came about as part of a discussion with Yuval Filmus and Artur Riazanov; it complements a result in [FLRS23] which handles small (sub-linear) Hamming weights. Our proof generalizes to weight m/h where h is not a power of 2. But it isn't clear how to handle even say $h = 4$. The general idea is as in [Vio12a]: each bit in S has $\mathbb{P}[1] = 1/3$, but decision trees have $\mathbb{P}[1] = a/2^q$ for an integer a. But the adaption to the sampling setting is not straightforward. We don't need the separator, but we make use of other lemmas in this paper to "find" a specific bit witnessing the difference in $\mathbb{P}[1]$.

Proof. Suppose $f(U)$ has statistical distance $\le 1 - \epsilon$ from S where f is a depth-q forest which uses s bits of input. We apply Lemma 15 to $P := f(U)$. Given $S_0 \subseteq S$ from the lemma, we let D be the preimage of S_0 according to f. By the lemma, $|D|/2^s \ge \Omega(\epsilon)$. Also, $H_\infty(f(D)) \ge H_\infty(S) - O(\log 1/\epsilon)$.

Now we use the fixed-set Lemma 19 with $\alpha := 2^{-q}/10$ and $w := 1$. Because D has loss $\log 1/\epsilon + O(1)$, the lemma gives us $D_1 \subseteq D$ with loss $2^{O(q)} \log 1/\epsilon$ s.t. D_1 is α-indistinguishable by depth-q trees from a product distribution D' in which every bit is either fixed or uniform. Hence any tree in $f(D_1)$ has $\mathbb{P}[1]$ which has distance $\le \alpha$ from $a/2^q$ for an integer a, because the probability is $a/2^q$ under D'.

However, $H_\infty(f(D_1)) \ge H_\infty(S) - O(\log 1/\epsilon) - 2^{O(q)} \log 1/\epsilon \ge H_\infty(S) - 2^{O(q)} \log 1/\epsilon$. This is $\ge H_\infty(S) - m^{1-\Omega(1)}$ for $q \le b \log m$ and $\epsilon \ge 1/2^{m^b}$ for suitable constant b.

To conclude the argument we have to show that any distribution supported within S with such large entropy has a bit with $\mathbb{P}[1]$ with distance $< 2^{-q}/10$ from $1/3$. This concludes the argument because it would mean that $1/3$ is within less than $2^{-q}/3$ from a number of the form $a/2^q$, which isn't true because $|a/2^q - 1/3| < 2^{-q}/3$ implies $|3a - 2^q| < 1$ which is false since 3 does not divide 2^q.

Because $S_0 \subseteq S$, the distribution $f(D_1)$ only contains strings of weight $m/3$. By the bound on the entropy, this distribution is a convex combination of distributions uniform over sets of size $|S|/2^{m^{1-\Omega(1)}}$ (see e.g. Lemma 6.10 in [Vad12]. Let X be any such distribution.

Let Y be the distribution on m i.i.d. bits with $\mathbb{P}[1] = 1/3$. S equals Y conditioned on an event of prob. $\ge \Omega(1/\sqrt{m})$. Hence X equals S conditioned on an event of prob. $\ge 2^{-m^{1-\Omega(1)}}$. A standard entropy argument such as Lemma 2.2 in [Vio12a] (which is a minor variant of Lemma 22 to account for non-uniform random variables) guarantees that all but $m - m^{1-\Omega(1)}$ bits have the same $\mathbb{P}[1]$ before or after the conditioning, up to an error of $1/m^{\Omega(1)}$. The result follows. $\square$

Acknowledgments. We thank the anonymous reviewers for detailed and helpful feedback.

References

[AB09] Sanjeev Arora and Boaz Barak. *Computational Complexity*. Cambridge University Press, 2009. A modern approach.

[Ajt83] Miklós Ajtai. Σ_1^1-formulae on finite structures. *Annals of Pure and Applied Logic*, 24(1):1–48, 1983.

[ASTS+03] Andris Ambainis, Leonard J. Schulman, Amnon Ta-Shma, Umesh V. Vazirani, and Avi Wigderson. The quantum communication complexity of sampling. *SIAM J. on Computing*, 32(6):1570–1585, 2003.

[BCS14] Itai Benjamini, Gil Cohen, and Igor Shinkar. Bi-lipschitz bijection between the boolean cube and the hamming ball. In *IEEE Symp. on Foundations of Computer Science (FOCS)*, 2014.

[BDT16] Avraham Ben-Aroya, Dean Doron, and Amnon Ta-Shma. Explicit two-source extractors for near-logarithmic min-entropy. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:88, 2016.

[BIL12a] Chris Beck, Russell Impagliazzo, and Shachar Lovett. Large deviation bounds for decision trees and sampling lower bounds for AC0-circuits. *Electronic Colloquium on Computational Complexity (ECCC)*, 19:42, 2012.

[BIL12b] Chris Beck, Russell Impagliazzo, and Shachar Lovett. Large deviation bounds for decision trees and sampling lower bounds for AC0-circuits. In *IEEE Symp. on Foundations of Computer Science (FOCS)*, pages 101–110, 2012.

[BIS12] Paul Beame, Russell Impagliazzo, and Srikanth Srinivasan. Approximating ac^0 by small height decision trees and a deterministic algorithm for #ac^0sat. In *Proceedings of the 27th Conference on Computational Complexity, CCC 2012, Porto, Portugal, June 26-29, 2012*, pages 117–125. IEEE Computer Society, 2012.

[CGZ21] Eshan Chattopadhyay, Jesse Goodman, and David Zuckerman. The space complexity of sampling. *Electron. Colloquium Comput. Complex.*, page 106, 2021.

[CK82] Imre Csiszar and Janos Korner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, Inc., 1982.

[Coh16] Gil Cohen. Making the most of advice: New correlation breakers and their applications. In *IEEE Symp. on Foundations of Computer Science (FOCS)*, pages 188–196, 2016.

[CS16] Gil Cohen and Leonard J. Schulman. Extractors for near logarithmic min-entropy. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:14, 2016.

[CT06] Thomas Cover and Joy Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

[CZ16] Eshan Chattopadhyay and David Zuckerman. Explicit two-source extractors and resilient functions. In *ACM Symp. on the Theory of Computing (STOC)*, pages 670–683, 2016.

[DW11] Anindya De and Thomas Watson. Extractors and lower bounds for locally samplable sources. In *Workshop on Randomization and Computation (RANDOM)*, 2011.

[ER60] P. Erdős and R. Rado. Intersection theorems for systems of sets. *J. London Math. Soc.*, 35:85–90, 1960.

[FLRS23] Yuval Filmus, Itai Leigh, Artur Riazanov, and Dmitry Sokolov. Sampling and certifying symmetric functions. In *APPROX-RANDOM*, 2023.

[FSS84] Merrick L. Furst, James B. Saxe, and Michael Sipser. Parity, circuits, and the polynomial-time hierarchy. *Mathematical Systems Theory*, 17(1):13–27, 1984.

[GM07] Anna Gál and Peter Bro Miltersen. The cell probe complexity of succinct data structures. *Theoretical Computer Science*, 379(3):405–417, 2007.

[Gol09] Alexander Golynski. Cell probe lower bounds for succinct data structures. In *20th ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 625–634, 2009.

[GSV18] Aryeh Grinberg, Ronen Shaltiel, and Emanuele Viola. Indistinguishability by adaptive procedures with advice, and lower bounds on hardness amplification proofs. In *IEEE Symp. on Foundations of Computer Science (FOCS)*, 2018. Available at http://www.ccs.neu.edu/home/viola/.

[GW20] Mika Göös and Thomas Watson. A lower bound for sampling disjoint sets. *ACM Trans. Comput. Theory*, 12(3):20:1–20:13, 2020.

[Hås87] Johan Håstad. *Computational limitations of small-depth circuits*. MIT Press, 1987.

[Hås14] Johan Håstad. On the correlation of parity and small-depth circuits. *SIAM J. on Computing*, 43(5):1699–1708, 2014.

[IMP12] Russell Impagliazzo, William Matthews, and Ramamohan Paturi. A satisfiability algorithm for AC^0. In *ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 961–972, 2012.

[Juk12] Stasys Jukna. *Boolean Function Complexity: Advances and Frontiers*. Springer, 2012.

[Li16] Xin Li. Improved two-source extractors, and affine extractors for polylogarithmic entropy. In *IEEE Symp. on Foundations of Computer Science (FOCS)*, 2016.

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

[LY20] Mingmou Liu and Huacheng Yu. Lower bound for succinct range minimum query. In Konstantin Makarychev, Yury Makarychev, Madhur Tulsiani, Gautam Kamath, and Julia Chuzhoy, editors, *ACM Symp. on the Theory of Computing (STOC)*, pages 1402–1415. ACM, 2020.

[Mil99] Peter Bro Miltersen. Cell probe complexity - a survey, 1999. Invited talk/paper at Advances in Data Structures (Pre-conference workshop of FSTTCS'99).

[O'D14] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.

[Păt08] Mihai Pătraşcu. Succincter. In *49th IEEE Symp. on Foundations of Computer Science (FOCS)*. IEEE, 2008.

[PT06] Mihai Patrascu and Mikkel Thorup. Time-space trade-offs for predecessor search. In Jon M. Kleinberg, editor, *ACM Symp. on the Theory of Computing (STOC)*, pages 232–240. ACM, 2006.

[PV10] Mihai Pătraşcu and Emanuele Viola. Cell-probe lower bounds for succinct partial sums. In *21th ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 117–122, 2010.

[Raz15] Alexander A. Razborov. Pseudorandom generators hard for k-DNF resolution and polynomial calculus resolution. *Ann. of Math.*, 181(2):415–472, 2015.

[SBI04] Nathan Segerlind, Sam Buss, and Russell Impagliazzo. A switching lemma for small restrictions and lower bounds for k-DNF resolution. *SIAM J. on Computing*, 33(5):1171–1200, 2004.

[Tho13] Mikkel Thorup. Mihai patrascu: Obituary and open problems. *Bulletin of the EATCS*, 109:7–13, 2013.

[Vad12] Salil P. Vadhan. Pseudorandomness. *Foundations and Trends in Theoretical Computer Science*, 7(1-3):1–336, 2012.

[Vio] Emanuele Viola. The Complexity of Distributions, Fall 2018 talk at the Simons Institute. https://www.youtube.com/watch?v=O78b085HE3w.

[Vio12a] Emanuele Viola. Bit-probe lower bounds for succinct data structures. *SIAM J. on Computing*, 41(6):1593–1604, 2012.

[Vio12b] Emanuele Viola. The complexity of distributions. *SIAM J. on Computing*, 41(1):191–218, 2012.

[Vio12c] Emanuele Viola. Extractors for turing-machine sources. In *Workshop on Randomization and Computation (RANDOM)*, 2012.

[Vio14] Emanuele Viola. Extractors for circuit sources. *SIAM J. on Computing*, 43(2):355–972, 2014.

[Vio19] Emanuele Viola. Lower bounds for data structures with space close to maximum imply circuit lower bounds. *Theory of Computing*, 15:1–9, 2019. Available at http://www.ccs.neu.edu/home/viola/.

[Vio20] Emanuele Viola. Sampling lower bounds: boolean average-case and permutations. *SIAM J. on Computing*, 49(1), 2020. Available at http://www.ccs.neu.edu/home/viola/.

[Vio22] Emanuele Viola, 2022. https://emanueleviola.wordpress.com/2022/09/14/myth-creation-the-switching-lemma/.

[Vio23] Emanuele Viola. New sampling lower bounds via the separator. In *Conf. on Computational Complexity (CCC)*, 2023. Available at http://www.ccs.neu.edu/home/viola/.

[Yao81] Andrew Chi-Chih Yao. Should tables be sorted? *J. of the ACM*, 28(3):615–628, 1981.

[Yao85] Andrew Yao. Separating the polynomial-time hierarchy by oracles. In *26th IEEE Symp. on Foundations of Computer Science (FOCS)*, pages 1–10, 1985.

[Yu19] Huacheng Yu. Optimal succinct rank data structure via approximate nonnegative tensor decomposition. In Moses Charikar and Edith Cohen, editors, *ACM Symp. on the Theory of Computing (STOC)*, pages 955–966. ACM, 2019.

A Proof of the fixed-set Lemma 19

Begin with R equal to the uniform distribution over W^s. If there are q words and q values such that the probability of getting those values in B is larger than $(1+\alpha)/W^q$ then we fix them to those values, in both B and R. Now we have subsets of W^{s-q}, the loss has decreased by an additive $\log_2 1/(1+\alpha) = \Omega(\alpha)$, and we repeat the process.

Because the initial loss was b, this process stops after $O(b/\alpha)$ iterations. In the end, the loss inside the final universe is at most b, since we never increase loss. With respect to the original universe, because we fixed $O(qb/\alpha)$ words, the loss is at most $O(wqb/\alpha) + b \le b \cdot O(wq/\alpha)$.

Let B_1 and R be the distributions when the process stops. Consider any tree $g : W^s \to \{0,1\}$ of depth q. Let P be the collection of paths in g leading to the output 1. Note that each path $p \in P$ corresponds to q input words and q values for them. Write $\mathbb{P}_X[p]$ for the probability of following path p under distribution X. By above we have $\mathbb{P}_{B_1}[p] \leq (1+\alpha)/W^q = (1+\alpha)\mathbb{P}_R[p]$.

Hence

$$\mathbb{P}[g(B_1) = 1] = \sum_{p \in P} \mathbb{P}_{B_1}[p] \leq \sum_{p \in P} (1+\alpha)\mathbb{P}_R[p] = (1+\alpha)\mathbb{P}[g(R) = 1].$$

And so in particular $\mathbb{P}[g(B_1) = 1] \leq \mathbb{P}[g(R) = 1] + \alpha$.

Repeating the argument with 0 and 1 swapped yields the lemma for trees with boolean alphabet. To prove the lemma for a tree g' with arbitrary alphabet, reduce to the case of boolean alphabet in the following standard way. Suppose that the statistical distance between $g'(R)$ and $g'(B_1)$ is $> \alpha$. This means that there exists a set T such that

$$|\mathbb{P}[g'(R) \in T] - \mathbb{P}[g'(B_1) \in T]| > \alpha.$$

Define tree g with boolean output as $g(x) := 1$ iff $g'(x) \in T$; note this just amounts to changing the labels of the leaves of g'. Now the left-hand side of the inequality above can be written as

$$|\mathbb{P}[g(R) = 1] - \mathbb{P}[g(B_1) = 1]|$$

and this contradicts the result for trees with boolean outputs and concludes the proof of the lemma.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il