Electronic Colloquium on Computational Complexity, Report No. 166 (2017)

A sampling lower bound for permutations

Emanuele Viola

October 17, 2017

Abstract

A map $f : [n]^\ell \to [n]^n$ has locality d if each output symbol in $[n] = \{1, 2, \ldots, n\}$ depends only on d of the ℓ input symbols in $[n]$. We show that the output distribution of a d-local map has statistical distance at least $1 - 2 \cdot \exp(-n/\log^{c^d} n)$ from a uniform permutation of $[n]$. This seems to be the first lower bound for the well-studied problem of generating permutations. Because poly(n)-size AC^0 circuits can sample almost uniform permutations, the result also separates AC^0 sampling from local.

As an application, we prove that any cell-probe data structure for storing permutations π of n elements such that $\pi(i)$ can be retrieved with d non-adaptive probes must use space $\geq \log_2 n! + n/\log^{c^d} n$. This is arguably the first lower bound for a natural data structure problem that is obtained from a sampling lower bound.

1 Introduction, results, and discussion

Permutations are fundamental objects which permeate computer science. A large literature has studied several problems on permutations, such as fast generation of a nearly uniform permutation and data structures for permutations.

Generating nearly-uniform permutations efficiently. There exist several surprising algorithms to generate permutations efficiently. Matias and Vishkin [MV91] and Hagerup [Hag91] show that poly(n)-size AC^0 circuits can generate a uniform random permutation of $[n] = \{1, 2, \ldots, n\}$, up to an exponentially small statistical error. (Their context is slightly different, for a streamlined presentation of the said result see [Vio12a].) They give several algorithmic applications of this result, and more applications have been found since then. The latter include sampling the output of any symmetric function together with its input in AC^0 [Vio12a], and constructing efficient secret-sharing schemes [BIVW16].

Another line of works studies generating random permutations using *switching networks*. A recent paper by Czumaj [Czu15] gives an explicit construction of switching networks with depth $O(\log^2 n)$ and $O(n \log n)$ switches that generate a nearly-uniform permutation on n elements, improving on previous work (see [Czu15] for discussion). The paper also conjectures that the depth can be improved to $O(\log n)$, and proves a partial result in this direction.

ISSN 1433-8092

On the side of lower bounds apparently nothing was known, and the above algorithms and conjectures arguably explain the difficulty of proving negative results. In this paper we prove a lower bound in the cell-probe model, with the restriction that all probes are non-adaptive. Specifically, we divide the memory in ℓ cells of $\log n$ bits (all logarithms in this paper are in base 2 unless otherwise noted), which are initialized uniformly at random. We consider algorithms that output n cells representing a function from $[n]$ to $[n]$ in the natural way. Each output cell only depends on a small number d of input cells. There is no restriction on the number ℓ of input cells the algorithm may use, though $\ell \leq dn$ without loss of generality.

Theorem 1. *Let $f : [n]^\ell \to [n]^n$ be a d-local map, i.e., a map such that each output symbol in $[n]$ depends only on d input symbols in $[n]$. Let $\Pi \in [n]^n$ be a random permutation of n elements. Let $f(U)$ be the output distribution of f for a uniformly chosen U in $[n]^\ell$. Then the statistical distance between $f(U)$ and Π is at least $1 - 2 \cdot \exp(-n/\log^{c^d} n)$ for an absolute constant c, where $\exp(x) = 2^x$.*

Theorem 1 remains nontrivial for locality d up to $d = \epsilon \log\log n$ for a small enough constant ϵ. (The factor 2 in the conclusion makes the bound trivially true if d is larger.) Note that the 1-local identity map $f(x) = x$ achieves statistical distance $1 - \exp(O(n))$, so for small locality the statistical bound in Theorem 1 is not far from optimal. Such a large statistical distance lower bound enables an application discussed next.

Data structures. The work [Vio12a] shows that sampling lower bounds with large statistical distance such as in Theorem 1 imply lower bounds for *static data structures.* This approach to obtain data structure lower bounds remains mostly unexplored. Although some data structure lower bounds proved this way appear in [Vio12a, LV12, BIL12], these bounds are either very weak or concern unnatural data structure problems. We suggest that the sampling approach could be used to attack some of the long-standing open problems in the area. Two central open problems are improving Siegel's state-of-the-art 1989 lower bound (Theorem 3.1 in [Sie04], rediscovered in [Lar12]), or proving lower bounds for the succinct dictionary problem (Problem 5 in Patrascu's obituary [Tho13]).

In this paper we obtain arguably the first new, natural data structure lower bound that is obtained from a sampling lower bound. The data structure problem is that of storing a permutation $\pi : [n] \to [n]$ so that $\pi(i)$ can be retrieved fast. At one extreme one can use $n \log_2 n$ bits to store the permutation and answer each query $\pi(i)$ by reading just one cell of $\log_2 n$ bits, at the other extreme we can use the minimum amount $\lceil \log_2 n! \rceil$ of space and answer queries by reading the entire memory. Our goal is to understand what is the right tradeoff between the time it takes to answer a query and *redundancy*, the amount of extra space the data structure needs over $\lceil \log_2 n! \rceil$. As a corollary to Theorem 1 we obtain the following tradeoff.

Corollary 2. *Consider any cell-probe data structure for storing permutations π of n elements such that $\pi(i)$ can be retrieved with d non-adaptive probes in cells of $\log_2 n$ bits. The data structure must use space $\geq \log_2 n! + n/\log^{c^d} n$ bits.*

Proof. If we fill the memory uniformly at random, we will be uniform over encodings of permutations with probability 2^{-r}. Hence if we run the data structure algorithm on uniform memory we obtain a sampler with statistical distance $< 1 - 2^{-r}$. The result then follows from Theorem 1. □

In particular, for constant time $d = O(1)$ the redundancy is $r \geq n/\text{poly}\log n$. By contrast, for other important problems there are surprising data structures that can achieve $d = O(1)$ and $r = O(1)$, and are also non-adaptive [Păt08, DPT10] (for an exposition of the relevant result in [DPT10] see [Vio09], Lectures 23-24).

Previous work has studied the problem of storing π so that $\pi(i)$ *and both* $\pi^{-1}(i)$ can be retrieved fast. [MRRR11] give several data structures for this problem. In particular, they give a data structure that can store a permutation using $\log_2 n! + n/\log^{2-o(1)} n$ bits such that $\pi(i)$ (and both $\pi^{-1}(i)$) can be computed in time $O(\log n)/\log\log n$. This data structure is based on a switching network known as the Benes network. They achieve their saving by "brute-forcing" certain small components.

On the side of lower bounds, Golynski shows in [Gol09] that any cell-probe data structure for representing a permutation $\pi : [n] \to [n]$ so that π can be computed with t cell probes and π^{-1} with t' must use $\log_2 n! + \Omega(n \log n)/(t \cdot t')$. This bound essentially matches the data structure in [MRRR11] for $t = \log n$, but tight bounds are not known in other parameter regimes. His technique is unlikely to apply to our simpler problem where we do not have the inverse queries $\pi^{-1}(i)$. In fact, none of the available techniques seems applicable for this problem: essentially, the only other technique available is the one in [PV10]. That technique requires that the mutual information between two sets of t queries is $\Omega(t)$. However, a calculation shows that in the case of permutations the mutual information is at most $O(t(t/n))$, and this prevents us from obtaining any non-trivial bound reasoning as in [PV10].

The complexity of distributions. Theorem 1 contributes a new type of lower bounds to the growing literature on lower bounds for sampling tasks [Vio12a, LV12, Vio14, DW11, Vio12b, BIL12, BCS14, Vio16]. Previously, lower bounds with statistical distance approaching 1 exponentially fast were only known for the problem of sampling error-correcting codes. These lower bounds applied to AC^0 samplers. By contrast, as mentioned at the beginning of this paper, AC^0 can sample nearly uniform permutations. For tasks that can be sampled in AC^0, the previous sampling lower bounds were much weaker. Thus, this work gives a new, strong separation between the sampling power of AC^0 and small-locality maps.

The role of adaptivity. We emphasize that the long-standing data-structure open problems mentioned earlier are also open for non-adaptive probes, and, as also mentioned earlier, some of the best-known data structures are non-adaptive. On the other hand, the succinct data structure for permutations in [MRRR11] does use adaptivity to follow a path in a switching network.

So far, in the case of *static* data structure lower bounds (which is our interest here), the restriction to non-adaptive probes has not made it easier to obtain lower bounds. This is in stark contrast with *dynamic* data structure problems, where for non-adaptive data structures we have polynomial lower bounds [BL15], wheres for non-adaptive data structures the best available lower bounds remain polylogarithmic (for the state-of-the-art see [LWY17]).

Open problems Despite a lot of effort, we have not been able to extend Theorem 1 to the case of *adaptive* probes.

Challenge: Let $f : [n]^\ell \to [n]^n$ be a map such that each output symbol depends on $d = O(1)$ adaptively chosen input cells. Show that the output distribution of f has statistical distance $\Omega(1)$ from a uniform random permutation.

Another question is to separate the power of adaptive and non-adaptive cell probes. Consider the distribution D over $[n]^{R+n}$ where the first R cells D_i are uniform in $[n]^R$ and each other cell is sampled as follows: Pick a uniform, independent index j in $\{1, 2, \ldots, R\}$ and output D_j. By definition D can be sampled with 2 adaptive probes. We conjecture that for $R = n^{1-\Omega(1)}$ sampling D requires a large number of non-adaptive probes.

Finally, recall that the statistical bound in Theorem 1 is not far from optimal for small locality. At the other end of the spectrum, it is an interesting question what is the minimum locality sufficient to reduce the statistical distance to $1 - \Omega(1)$.

1.1 Techniques

Theorem 1 is proved by induction on the locality d. Consider a d-local map f and write $f = (f_1, f_2, \ldots, f_n)$ where f_i is the function outputting the cell i. In the induction step, we start with a relatively standard *covering argument.* That says that either we have (A) a small number of input cells C that intersect the probes made by all the f_i, or else (B) we have many f_i whose set of probes are disjoint.

In case (A), suppose we fix the contents of the cells C. Because every f_i probes a cell in C, this reduces the locality of f. Thus, we can write our sampler f as a convex combination of samplers with smaller locality, one for each possible fixing of the contents of the cells in C. To analyze this step we show (Corollary 5) that if D is a distribution that is a convex combination of 2^s distributions D_i (the samplers obtained by any possible fixing of the $s = |C| \log n$ bits in the cells C) where each D_i has statistical distance $\geq 1 - \epsilon$ from a target distribution T, then D has distance $\geq 1 - c^s \epsilon$ from T. By setting the parameters appropriately, we can ensure that $\epsilon \ll 1/c^s$, concluding this case.

In case (B), we have many f_i which are independent. We obtain large statistical distance just considering these independent f_i. The high-level idea is that if the f_i have small entropy, then the result follows because a uniform permutation has large entropy. Otherwise, if the f_i have high entropy we can show by the *birthday paradox* that the outputs of the f_i will collide (i.e., $f_i = f_j$ for some $i \neq j$) with high probability. Since this never happens for permutations, we obtain statistical distance.

Formalizing case (B) requires finding the right notion of “high-entropy.” If we have t independent f_i, we define one f_i to be “high-entropy” if for every set S of $t/2$ values, the probability that $f_i \in S$ is $\Omega(|S|/n)$. Now, if there are $t/2$ functions f_i that have high-entropy, then we can run a folklore, simplified proof of the birthday paradox: fix the other

$t/2$ functions arbitrarily, and define S to be the set of values they take. By high-entropy and independence, the probability of not having a collision will be

$$(1 - \Omega(|S|/n))^{t/2} \leq e^{-\Omega(t^2/n)}$$

which is small enough when $t = n^{0.5+\Omega(1)}$. Since a uniform permutation by definition never has a collision, we obtain statistical distance $1 - e^{-\Omega(t^2/n)}$.

If on the other hand we have $t/2$ functions which are low entropy, we use concentration of measure to show that they will land in their sets S too often. Here again we obtain a statistical distance $1 - e^{-\Omega(t^2/n)}$.

We shall start with $t = \Omega(n)$ for $d = 0$, and then progressively update it via $t \to t^2/n$ from the above abounds. Losing along the way $\log n$ factors that arise from having cells of $\log n$ bits, this gives the bound in Theorem 1.

2 Proof of Theorem 1

In this section we prove Theorem 1. First, in §2.1 we show that a convex combination of distributions that are distant from a target distribution remains distant. Then in §2.2 we show that any collection of independent random variables is distant from uniform variables conditioned on not colliding. Finally, in §2.3 we use these results to prove Theorem 1.

2.1 Combo of far distributions is far

We start with a lemma about two distributions and then we obtain our main result as a corollary.

Definition 3. The L_1 norm of a vector u is $L_1(u) := \sum_x |u(x)|$. The *statistical distance* between two vectors is $\Delta(u, v) := \frac{1}{2}L_1(u - v)$.

Lemma 4. *Let p and q and t be distributions over the same arbitrary domain. Let $r = \frac{1}{2}(p + q)$ be a convex combination of p and q. If $\Delta(p, t) \geq 1 - \epsilon$ and $\Delta(q, t) \geq 1 - \epsilon$ then $\Delta(r, t) \geq 1 - O(\epsilon)$. Moreover, there exist distributions for which the conclusion is $\Delta(r, t) = 1 - 2\epsilon$.*

Proof. First, we claim that without loss of generality we can assume that our distributions are over only $O(1)$ points. Indeed, if you have two points x and y in the domain where the order of p, q, r, t is the same (e.g., $p(x) > t(x) > r(x) > q(x)$, and the same for y) then you can sum the two points. This does not change the statistical distances (because all the absolute values have the same “sign”). Specifically, it clearly does not change the distances between t and p and q. To illustrate what happens to the distance between r and t, consider two points x and y where $r(x) > t(x)$ and $r(y) > t(y)$. Let’s put all the mass of x and y on a new point z. Then we have $r(z) = (p(x) + p(y) + q(x) + q(y))/2 = r(x) + r(y) > t(x) + t(y) = t(z)$. So the contribution to the statistical distance for z is $r(z) - t(z) = r(x) + r(y) - t(x) - t(y)$ which is the sum of the contributions for x and y.

Because there are only $O(1)$ choices for the order of p, q, r, t, we can have distributions over $O(1)$ points with all the distances preserved. Now we claim that on every point, if $t \geq \epsilon$ then $\max\{p, q\} \leq \epsilon$, and if $\max\{p, q\} \geq \epsilon$ then $t \leq \epsilon$. This is because otherwise the distances $\Delta(p, t), \Delta(q, t)$ are contradicted.

Now modify the distributions p, q, t by setting them to 0 if they are $\leq \epsilon$. This only changes their mass by $O(\epsilon)$, which is where we use that the support has size $O(1)$. Call the new vectors p', q', t'. Note these are non-negative vectors with sum (aka L_1 norm) in $[1 - O(\epsilon), 1]$. Let $r' = (p' + q')/2$, which has sum in $[1 - O(\epsilon), 1]$ too. Also,

$$\Delta(r, r') = \frac{1}{2}\sum_x |\frac{p(x) + q(x)}{2} - \frac{p'(x) + q'(x)}{2}| \leq \frac{1}{2}\sum_x |\frac{p(x) - p'(x)}{2}| + |\frac{q(x) - q'(x)}{2}|$$
$$= \frac{1}{2}(\Delta(p, p') + \Delta(q, q')) \leq O(\epsilon).$$

Now note $\Delta(r', t') \leq \Delta(r', r) + \Delta(r, t) + \Delta(t, t')$. So $\Delta(r, t) \geq \Delta(r', t') - O(\epsilon)$. There remains to prove a lower bound on $\Delta(r', t')$. But this is easy, for either $t' \geq 0$ or exclusively $r' \geq 0$. So $\Delta(t', r') = \frac{1}{2}\sum_x |t'(x) - r'(x)| = \frac{1}{2}\sum_x (t'(x) + r'(x)) \geq 1 - O(\epsilon)$, because recall the vectors have sum $1 - O(\epsilon)$.

To prove the last sentence in the lemma statement, consider the domain $\{1, 2, 3, 4\}$ and distributions as follows:
$p(1) = \epsilon, q(1) = 0, t(1) = \epsilon/2$,
$p(2) = 0, q(2) = \epsilon, t(2) = \epsilon/2$,
$p(3) = 0, q(3) = 0, t(3) = 1 - \epsilon$,
$p(4) = 1 - \epsilon, q(4) = 1 - \epsilon, t(4) = 0$.

Note that $r(i) = p(i) = q(i)$ for $i \in \{3, 4\}$. We have $\Delta(p, t) = \Delta(q, t) = \epsilon/2 + 1 - \epsilon = 1 - \epsilon/2$, but $\Delta(r, t) = 1 - \epsilon$. □

Corollary 5. *Let r and t be distributions over the same arbitrary domain. Suppose that $r = \frac{1}{2^s}\sum_{i=1}^{2^s} p_i$ and that each p_i is a distribution with $\Delta(p_i, t) \geq 1 - \epsilon$. Then $\Delta(r, t) \geq 1 - c^s\epsilon$, for a constant c.*

Proof. We proceed by induction on s. Let c be a constant so that Lemma 4 has conclusion $1 - c\epsilon$. Write $r = \frac{1}{2}(r_1 + r_2)$ where the r_i are convex combinations of 2^{s-1} distributions. By hypothesis $\Delta(r_1, t) \geq 1 - c^{s-1}\epsilon$, and the same holds for r_2. By Lemma 4, $\Delta(r, t) \geq 1 - c^s\epsilon$. □

2.2 Independent vs. permutation

We shall need a lemma about concentration of measure.

Lemma 6. *Let $x_1, x_2, \ldots, x_m$ be random variables such that for every i, conditioned on any outcome of all the variables except x_i, we have $\Pr[x_i = 1] \geq p$. For all sufficiently small p we have $\Pr[\sum x_i \leq 0.5pm] \leq \exp(-\Omega(pm))$.*

Similar lemmas have been proved many times. For completeness we give a proof relying on a bound in [PS97]. We use the presentation in [IK10].

Proof. Define $y_i := 1 - x_i$. We have $\Pr[y_i = 1] \leq 1 - p$ conditioned on any outcome of all the y variables except y_i. We need to bound $\Pr[\sum y_i \geq m(1-0.5p)]$. The variables y_i satisfy the property that for any set $S \subseteq [m]$, $\Pr[\forall i \in S, y_i = 1] \leq (1-p)^{|S|}$, because the probability can be written as $\Pr[y_{i_1} = 1] \cdot \Pr[y_{i_2} = 1 | y_{i_1} = 1] \cdots$, where $i_1, i_2, \ldots$ are the elements of S, and each term is at most $1 - p$. So we can apply Theorem 1.1 in [IK10] to obtain

$$\Pr[\sum y_i \geq m(1-0.5p)] \leq e^{-mD(1-0.5p|1-p)}$$

where D is the relative entropy defined as $D(x|y) = x\log_e(x/y) + (1-x)\log_e((1-x)/(1-y))$. From the definition we observe $D(x|y) = D(1-x|1-y)$, hence the above upper bound is $e^{-mD(0.5p|p)}$. Finally, we claim that $D(0.5p|p) \geq \Omega(p)$. Indeed, again by definition we have

$$D(0.5p|p) = 0.5p\log_e 0.5 + (1-0.5p)\log_e \frac{1-0.5p}{1-p}.$$

The first summand is about $0.5p \cdot (-0.693\ldots)$. The second can be written as $(1 - 0.5p)\log_e(1 + 0.5p/1-p)$. For small enough p this converges to $0.5p$. More precisely we have $\log_e(1+x) \geq x - x^2/2$ by Taylor approximation, and the result follows. □

We can now state and prove our main result of this subsection.

Lemma 7. *Let $x_1, x_2, \ldots, x_t$ be t independent random variables over $[n]$. Let Π be a random, uniform permutation over $[n]$. The statistical distribution between the x_i and $\Pi(1), \Pi(2), \ldots, \Pi(t)$ is at least $1 - \exp(-\Omega(t^2/n))$.*

Proof. Let $p := 0.5t/n$. Call a variable x_i *low-entropy* if there is a set S_i of size $t/2 = pn$ such that $\Pr[x_i \in S_i] \leq 0.1p$. We consider two cases:

Case 1: There are $t/2$ low-entropy variables x_i:

In this case select any $b := 0.1t$ low-entropy variables. Without loss of generality assume that they are $x_1, x_2, \ldots, x_b$ and let $Y_1, Y_2, \ldots, Y_b$ be the indicator variables corresponding to the events "$x_i \in S_i$". Consider the statistical test "$\sum_{i \leq b} Y_i \geq 0.2p \cdot b$". In the sampler case, $\mathbb{E}[\sum_{i \leq b} Y_i] \leq 0.1p \cdot b$. The probability that the test passes is at most the probability that $\sum Y_i$ deviates from its expectation by a constant factor. Without loss of generality we can assume that $\mathbb{E}[\sum_{i \leq b} Y_i]$ is exactly $0.1bp$. The variables are independent, and so by a Chernoff bound this probability is at most $\exp(-0.1bp) = \exp(-\Omega(t^2/n))$.

Now consider the permutation case and let $Y_1, Y_2, \ldots, Y_b$ be the indicator variables corresponding to the events "$\Pi(i) \in S_i$". We observe that regardless of the outcome of any other $r \leq b$ variables $\Pi(j)$, $j \neq i$, (note that $\Pi(j)$ determines Y_j)

$$\Pr[Y_i = 1] \geq \frac{|S| - r}{n - r} = \frac{0.5t - r}{n - r} \geq \frac{0.4t}{n} = 0.8p.$$

The probability that the test does not pass is at most the probability that $\sum_i Y_i < 0.5(0.8p)b$, and that by Lemma 6 is $\leq \exp(-\Omega(t/n) \cdot b) = \exp(-\Omega(t^2/n))$.

Case 2: There are not $t/2$ low-entropy variables x_i:

In this case there are $\geq t/2$ high-entropy variables, i.e., variables such that for every set S_i of size $t/2$, the probability of landing in S_i is $\geq 0.1p$. Let H be the index set of $t/2$ of these variables, and L be the index set of the other $t/2$ variables (which may or may not be high entropy). The probability that the x_i collide (i.e., two variables take the same value) is at least the probability that the variables collide conditioned on the event that there is no collision among the variables in L. Fix any outcome for the variables in L conditioned on the event that they do not collide. Because they do not collide, they take $t/2$ distinct values. Let S be the set of $t/2$ values they take. Now the probability that the x_i variables collide is at least the probability that some variable in H lands in S. Because the variables are independent, this probability is at least

$$1-(1-0.1p)^{t/2} \geq 1-e^{-\Omega(pt)} = 1-e^{-\Omega(t^2/n)}.$$

On the other hand, by definition, the variables $\Pi(i)$ never collide. Hence the statistical test that simply checks if the variables collide gives the desired statistical distance. □

2.3 Proof of Theorem 1

We proceed by induction on d. We can take $d=0$ as base case. In this case f is constant and the statistical distance is $1-1/n!$ which is larger than $1-2^{-n/\log n}$.

For the induction step, you ask the question whether there are

$$t := n/\log^{c^d/4} n$$

variables with indexes $T \subseteq [n]$ whose probes intersect the probes of all other variables. If the answer is affirmative, then by considering any possible fixing for the values of the cells probed by the variables in T your distribution is a convex combination of $2^{t\cdot d\cdot \log n}$ distributions which are $(d-1)$-local. By the induction hypothesis applied to each of these samplers, and Corollary 5 the statistical distance will be

$$1-O(1)^{t\cdot d\cdot \log n}\cdot 2^{-n/\log^{c^{d-1}} n}.$$

This quantity equals $1-2^{-x}$ where

$$x = \frac{n}{\log^{c^{d-1}} n} - O(td\log n) = n\left(\frac{1}{\log^{c^{d-1}} n} - \frac{O(d\log n)}{\log^{c^d/4} n}\right) \geq 0.5\frac{n}{\log^{c^{d-1}} n} \geq \frac{n}{\log^{c^d} n}.$$

Here the inequalities hold for $d \leq \log n$ say (for else the theorem is trivial) and a suitable choice of c.

If the answer is no then there are t variables which are independent. (This can be shown by iteratively collecting variables whose probes are disjoint. We can't stop before we collect t, for else the answer is yes.) By Lemma 7 just considering those variables the statistical distance is at least $1-\exp(-\Omega(t^2/n))$. Noting that $t^2/n = n/\log^{c^d/2} n \geq (\log^{c^d/2} n)n/\log^{c^d} n$ concludes the argument for all large enough n.

References

[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.

[BIL12] 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.

[BIVW16] Andrej Bogdanov, Yuval Ishai, Emanuele Viola, and Christopher Williamson. Bounded indistinguishability and the complexity of recovering secrets. In *Int. Cryptology Conf. (CRYPTO)*, 2016.

[BL15] Joshua Brody and Kasper Green Larsen. Adapt or die: Polynomial lower bounds for non-adaptive dynamic data structures. *Theory of Computing*, 11:471–489, 2015.

[Czu15] Artur Czumaj. Random permutations using switching networks. In *ACM Symp. on the Theory of Computing (STOC)*, pages 703–712, 2015.

[DPT10] Yevgeniy Dodis, Mihai Pătraşcu, and Mikkel Thorup. Changing base without losing space. In *42nd ACM Symp. on the Theory of Computing (STOC)*, pages 593–602. ACM, 2010.

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

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

[Hag91] Torben Hagerup. Fast parallel generation of random permutations. In *18th Coll. on Automata, Languages and Programming (ICALP)*, pages 405–416. Springer, 1991.

[IK10] Russell Impagliazzo and Valentine Kabanets. Constructive proofs of concentration bounds. In *Workshop on Randomization and Computation (RANDOM)*, pages 617–631. Springer, 2010.

[Lar12] Kasper Green Larsen. The cell probe complexity of dynamic range counting. In *ACM Symp. on the Theory of Computing (STOC)*, pages 85–94, 2012.

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

[LWY17] Kasper Green Larsen, Omri Weinstein, and Huacheng Yu. Crossing the logarithmic barrier for dynamic boolean data structure lower bounds. *CoRR*, abs/1703.03575, 2017.

[MRRR11] J. Ian Munro, Rajeev Raman, Venkatesh Raman, and S. Srinivasa Rao. Succinct representations of permutations and functions. *CoRR*, abs/1108.1983, 2011.

[MV91] Yossi Matias and Uzi Vishkin. Converting high probability into nearly-constant time-with applications to parallel hashing. In *23rd ACM Symp. on the Theory of Computing (STOC)*, pages 307–316, 1991.

[Păt08] Mihai Pătraşcu. Succincter. In *49th IEEE Symp. on Foundations of Computer*

Science (FOCS). IEEE, 2008.

[PS97] Alessandro Panconesi and Aravind Srinivasan. Randomized distributed edge coloring via an extension of the chernoff-hoeffding bounds. *SIAM J. Comput.*, 26(2):350–368, 1997.

[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.

[Sie04] Alan Siegel. On universal classes of extremely random constant-time hash functions. *SIAM J. on Computing*, 33(3):505–543, 2004.

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

[Vio09] Emanuele Viola. Gems of theoretical computer science. Lecture notes of the class taught at Northeastern University. Available at http://www.ccs.neu.edu/home/viola/classes/gems-08/index.html, 2009.

[Vio12a] Emanuele Viola. The complexity of distributions. *SIAM J. on Computing*, 41(1):191–218, 2012.

[Vio12b] 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.

[Vio16] Emanuele Viola. Quadratic maps are hard to sample. *ACM Trans. Computation Theory*, 8(4), 2016.