# Average Case Analysis of Brzozowski's Algorithm

Sven de Felice, Cyril Nicaud

International Journal of Foundations of Computer Science


# Average case analysis of Brzozowski's algorithm*

SVEN DE FELICE
*LIAFA, Université Paris Diderot - Paris 7 & CNRS UMR 7089, F-75205, Paris, France*
*Sven.De-Felice@liafa.univ-paris-diderot.f*

CYRIL NICAUD
*Université Paris-Est, LIGM (UMR 8049), UPEMLV, F-77454, Marne-la-Vallée, France*
*cyril.nicaud@univ-mlv.fr*

Received (Day Month Year)
Accepted (Day Month Year)
Communicated by (xxxxxxxxxx)

We analyze the average complexity of Brzozowski's minimization algorithm for distributions of deterministic automata with a small number of final states. We show that, as in the case of the uniform distribution, the average complexity is super-polynomial even if we consider random deterministic automata with only one final state. Such results were only known for distributions where the expected number of final states was linear in the number of states.

*Keywords*: automata minimization; Brzozowski's algorithm; average case analysis.

## 1. Introduction

In this article we analyze the average case complexity of Brzozowski's state minimization algorithm [3]. Recall that this method is based on the fact that determinizing a trim co-deterministic automaton that recognizes a language $\mathcal{L}$ yields the minimal automaton of $\mathcal{L}$. Hence, starting from an automaton $\mathcal{A}$ that recognizes the language $\mathcal{L}$, one can compute its minimal automaton by first determinizing its reversal, then by determinizing the reversal of the resulting automaton.

This elegant method is not efficient in the worst case, since the first determinization can produce an automaton that has exponentially many states, even if one starts with a deterministic automaton (see [8] for a classical example). We are therefore far from the efficient solutions available to minimize deterministic automata, such as Hopcroft's algorithm [12], which runs in $\mathcal{O}(n \log n)$ time.

In [8] we proved that for the uniform distribution on deterministic and complete automata with $n$ states, or for distributions where each state is final with

*This work is supported by the French National Agency (ANR) through ANR-2010-BLAN-0204.

(fixed) probability $b \in (0,1)$, the running time of Brzozowski's algorithm is super-polynomial[a] with high probability. For such distributions, on average, an asymptotically constant proportion of the states are final. In many situations, however, the automata considered have few final states. Thus, a natural question is whether this result still holds for automata with, for instance, a fixed number of final states. This is the question we investigate in this article.

Our setting is defined precisely in Section 2.2. It covers the cases of random automata with just one final state, with $\log n$ final states, or where each state is final with probability $\frac{3}{n}$ or $\frac{2}{\sqrt{n}}$, and so on. It therefore differs significantly from the cases studied in [8].

The analysis of distributions of automata with a small number of final states is an active topic in the statistical study of automata (see [14] for a recent survey). The main results in this field, the average complexity of Moore's algorithm and the asymptotic number of minimal automata, only hold for distributions of automata with "sufficiently many" final states [1, 6, 2]. Efforts have been undertaken to extend them to, say, automata with only one final state, but with no success so far. To our knowledge, we present in this article the first result of this kind.

The paper is organized as follows. After recalling some basic notions and defining our distributions on automata in Section 2, we state our main result in Section 3. Our principal tool from automata theory is then presented in Section 4. Section 5 is devoted to the proof of our main theorem. Some further directions are finally proposed in Section 6.

This article is the full version of the extended abstract [9], where all the missing proofs have been included.

## 2. Definitions

Let $[n]$ denote the set of integers between 1 and $n$. If $x, y$ are two real numbers, let $[\![x, y]\!]$ denote the set of integers $i$ such that $x \leq i \leq y$. For any positive integer $n$, let $\mathfrak{S}_n$ denote the set of all permutations on $[n]$.

### 2.1. *Automata*

An *automaton* $\mathcal{A}$ is a tuple $(A, Q, \delta, I, F)$, where $A$ is its non-empty finite alphabet, $Q$ is its finite set of *states*, $I \subseteq Q$ is its set of *initial states* and $F \subseteq Q$ is its set of *final states*. Its *transition function* $\delta$ is a (possibly partial) map from $Q \times A$ to $2^Q$. A *transition* of $\mathcal{A}$ is a tuple $(p, a, q) \in Q \times A \times Q$, which we write $p \xrightarrow{a} q$, such that $q \in \delta(p, a)$. The map $\delta$ is classically extended by morphism to $Q \times A^*$. We denote by $\mathcal{L}(\mathcal{A})$ the set of words recognized by $\mathcal{A}$. A *deterministic and complete automaton* is an automaton such that $|I| = 1$ and for every $p \in Q$ and $a \in A$, $|\delta(p, a)| = 1$; for such an automaton, we consider $\delta$ as a total map from $Q \times A^*$ to

[a] A sequence is *super-polynomial* when it grows faster than $n^\gamma$, for any positive $\gamma$.

$Q$ to simplify the notation. A state $q$ is *accessible* when there exists a path from an initial state to $q$. It is *co-accessible* when there exists a path from $q$ to a final state. An automaton is *trim* when all its states are accessible and co-accessible. If $\mathcal{A}$ is an automaton, let $\mathrm{Trim}(\mathcal{A})$ denote the automaton obtained after removing states that are not accessible or not co-accessible.

For any automaton $\mathcal{A} = (A, Q, \delta, I, F)$, let $\tilde{\mathcal{A}}$ be the *reverse* of $\mathcal{A}$. It is the automaton $\tilde{\mathcal{A}} = (A, Q, \tilde{\delta}, F, I)$, where $p \xrightarrow{a} q$ is a transition of $\tilde{\mathcal{A}}$ if and only if $q \xrightarrow{a} p$ is a transition of $\mathcal{A}$. The automaton $\tilde{\mathcal{A}}$ recognizes the reverse[b] of $\mathcal{L}(\mathcal{A})$. An automaton is *co-deterministic* when its reverse is deterministic.

Recall that the *minimal automaton* of a rational language $\mathcal{L}$ is the smallest deterministic and complete automaton[c] that recognizes $\mathcal{L}$. To each rational language $\mathcal{L}$ corresponds a minimal automaton, which is unique up to isomorphism. The *state complexity* of a regular language is the number of states of its minimal automaton.

If $\mathcal{A} = (A, Q, \delta, I, F)$ is a non-deterministic automaton, the subset automaton of $\mathcal{A}$ is the automaton $\mathcal{B} = \big(2^Q, \delta', \{I\}, \{X \in 2^Q \mid F \cap X \neq \emptyset\}\big)$, where $\delta'(X, a) = \cup_{p \in X} \delta(p, a)$, for every $X \subseteq Q$ and every $a \in A$. It is a deterministic automaton that recognizes the same language as $\mathcal{A}$. This is still true if we only take the accessible part of $\mathcal{B}$. This accessible part can be built on the fly, using the rule for $\delta'$ in any traversal of $\mathcal{B}$, starting from $I$. We denote by $\mathrm{Subset}(\mathcal{A})$ the accessible part of the subset automaton of $\mathcal{A}$.

In [3], Brzozowski established the following result:

**Theorem 1 (Brzozowski)** *If $\mathcal{A}$ is a trim co-deterministic automaton then Subset($\mathcal{A}$) is the minimal automaton of $\mathcal{L}(\mathcal{A})$.*

This theorem readily yields an algorithm to compute the minimal automaton of $\mathcal{L}(\mathcal{A})$, based on the subset construction: since $\mathcal{B} = \mathrm{Subset}(\mathrm{Trim}(\tilde{\mathcal{A}}))$ is a deterministic automaton recognizing the reverse of $\mathcal{L}(\mathcal{A})$, then $\mathrm{Subset}(\mathrm{Trim}(\tilde{\mathcal{B}}))$ is the minimal automaton of $\mathcal{L}(\mathcal{A})$. This is known as Brzozowski's minimization algorithm.

### 2.2. *Probabilities on automata*

A *mapping of size $n$* is a total function from $[n]$ to $[n]$. A mapping $f$ can be seen as a directed graph with an edge $i \to j$ whenever $f(i) = j$. Such a graph is a union of cycles of Cayley trees (i.e., rooted labelled trees), as depicted in Fig. 1 (see [10] for more information on this graph description). Let $f$ be a mapping of size $n$. An element $x \in [n]$ is a *cyclic point* of $f$ if there exists an integer $i > 0$ such that $f^i(x) = x$. The *cyclic permutation* of a mapping $f$ is the permutation obtained when restricting $f$ to its set of cyclic points. The *normalized cyclic permutation* of $f$ is obtained by relabelling the $c$ cyclic points of $f$ with the elements of $[c]$, while

[b]If $u = u_0 \cdots u_{n-1}$ is a word of length $n$, the *reverse* of $u$ is the word $\tilde{u} = u_{n-1} \cdots u_0$.
[c]Minimal automata are not always required to be complete in the literature.

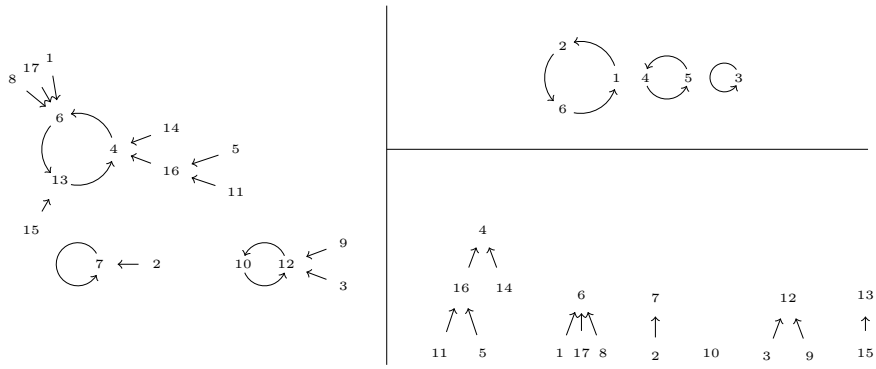

Fig. 1. A mapping of size 17, on the left. On the upper right we have its normalized cyclic part, and on the lower right its Cayley trees (not normalized).

keeping their relative order[d].

In the sequel, $A$ is always a fixed alphabet with $k \geq 2$ letters. Let $\mathfrak{A}_n$ denote the set of all deterministic and complete automata with input alphabet $A$ whose set of states is $[n]$ and whose initial state is 1. Such an automaton $\mathcal{A}$ is characterized by the tuple $(n, \delta, F)$. A *transition structure* is an automaton without final states. Let $\mathfrak{T}_n$ denote the set of transition structures with $n$ states and with the same rules of labelling we used for $\mathfrak{A}_n$. If we want to specify the alphabet $A$, we write $\mathfrak{T}_n(A)$ instead of $\mathfrak{T}_n$. If $\mathcal{A} \in \mathfrak{A}_n$ and $a \in A$, an $a$-cycle of $\mathcal{A}$ is a cycle of the mapping induced by $a$, i.e., $p \mapsto \delta(p, a)$. A state of an $a$-cycle is called an *$a$-cyclic state*.

Let $E$ be a set of combinatorial objects, equipped with a notion of size, such that the subset $E_n \subseteq E$ of elements of size $n$ is finite for every $n \geq 0$. The *uniform distribution* (which is a slight abuse of notation since there is one distribution for each $n$) on the set $E$ is defined for any $e \in E_n$ by $\mathbb{P}_n(\{e\}) = \frac{1}{|E_n|}$. The reader is referred to Part C of [11] for more information on combinatorial probabilistic models.

Consider a (sequence of) distributions on $E = \bigcup_n E_n$. Let $P$ be a property defined on $E$. We say that $P$ holds *with high probability* when a random object from $E_n$ has $P$ with probability that tends to 1 as $n$ tends to infinity. We say that $P$ is *visible* (or holds with *visible probability*) when there exists a constant $C > 0$ and an integer $n_0$ such that for every $n \geq n_0$, a random element of $E_n$ has $P$ with probability at least $C$.

For any given $n$, a distribution on the subsets of $[n]$ is *label-independent* if any two subsets of the same cardinality have the same probability.

We are interested in the distributions on $\mathfrak{A}_n$ such that the transition structure

[d]The notion of normalization will be used for other substructures, always for relabelling the atoms with an initial segment of the positive integers, while keeping their relative order.

is chosen uniformly in $\mathfrak{T}_n$, and the set of final state is chosen independently, using a label-independent distribution. We call such distributions *good*. If furthermore the number of final states is in $[\![1, \frac{n}{2}]\!]$ with visible probability, we say that the distribution is a *a good distribution with a small number of final states*.

Natural examples of good distributions with a small number of final states are:

- The *fixed-size distribution* on $\mathfrak{A}_n$ of parameter $m$, where the set of final states is chosen uniformly amongst all subsets of size $m$ of $[n]$, for $m \leq \frac{n}{2}$. The parameter $m$ may depend[e] on $n$; one can for instance consider the fixed-size distribution of parameter $\lfloor\sqrt{n}\rfloor$.
- The *p-distribution* on $\mathfrak{A}_n$, where each state is final with probability $p \leq \frac{1}{2}$, independently; in this model also, $p$ may depend on $n$, for instance $p = \frac{2}{n}$ yields automata with two final states on average.

Notice that for $p = \frac{1}{2}$, the $p$-distribution is the uniform distribution on $\mathfrak{A}_n$. The results presented in this article therefore covers the uniform case. However, they are weaker than those established in [9], which is dedicated to such distributions.

## 3. Main result

Our main result is the following:

**Theorem 2.** *Let $A$ be an alphabet with at least 3 letters. If $\mathcal{A}$ is a random deterministic and complete automaton with $n$ states following a good distribution with a small number of final states, then, for any positive $\gamma$, the state complexity of the reverse of $\mathcal{L}(\mathcal{A})$ is at least $n^\gamma$ with visible probability.*

Compared to the main result of [8], we capture many more distributions on automata, by weakening the statement: it does not hold with high probability but with positive probability only, and we need an alphabet with at least three letters. The result cannot hold with high probability: as proved in [4], there is a linear number of states that are not accessible in a typical random automaton; thus for the fixed size distribution with one final state, the final state has a positive probability of not being accessible. The result may hold for binary alphabets, see Section 6 for a discussion on that point. Notice also that we slightly improved over [9] by generalizing the possible distributions for the set of final states.

The average complexity of Brzozowski's algorithm is a direct consequence of Theorem 2.

**Corollary 3.** *Let $A$ be an alphabet with at least 3 letters. The average complexity of Brzozowski's algorithm is super-polynomial for good distributions with a small number of final states.*

[e]The term "fixed" stands for: for any given $n$, the number of final states is fixed.

**Proof.** For any $\gamma > 0$, the expected number of states after the first determinization is at least $n^\gamma$ times the probability that an automaton has at least $n^\gamma$ states after the first determinization. By Theorem 2, this probability is greater than some positive constant $C$ for $n$ sufficiently large, concluding the proof. □

## 4. A criterion to lower bound the state complexity of the reverse

Recall that $\mathfrak{A}_n$ only contains automata that are deterministic and complete.

Let $\mathcal{A} = (n, \delta, F)$ be an automaton of $\mathfrak{A}_n$, let $\alpha \in A$ and let $u \in A^*$. We define the *$u$-word* $v \in \{0,1\}^\ell$ of an $\alpha$-cycle $\mathcal{C}$ of length $\ell$ as follows. If $x$ is the smallest element of $\mathcal{C}$, we set $v_i = 1$ if and only if $\delta(x, \alpha^i u) \in F$, for $i \in \{0, \ldots, \ell - 1\}$. In other words, starting from $x$, we follow the cycle and record whether $u$ maps each state to a final state or not. The cycle $\mathcal{C}$ is *$u$-primitive* if its $u$-word $v$ is a primitive word (i.e., $v$ cannot be written as $v = w^t$ for some word $w$ and some $t \geq 2$).

**Lemma 4.** *Let $\alpha \in A$ and $u \in A^*$. Let $\mathcal{A} \in \mathfrak{A}_n$ be an automaton that contains $m$ distinct accessible $u$-primitive $\alpha$-cycles $\mathcal{C}_1, \ldots, \mathcal{C}_m$ of length at least two. Then the state complexity of $\mathcal{L}(\tilde{\mathcal{A}})$ is at least* $\operatorname{lcm}(|\mathcal{C}_1|, \ldots, |\mathcal{C}_m|)$.

**Proof.** Let $\mathcal{A} = (n, \delta, F)$. By Theorem 1, the minimal automaton of the reverse of $\mathcal{L}(\mathcal{A})$ is obtained by determinizing the trim part of the reverse of $\mathcal{A}$. Let $F' \subseteq F$ be the set of accessible final states, and let $U = \delta^{-1}(F', u)$. The set $F'$ is not empty by hypothesis on the $\mathcal{C}_i$'s. Clearly $U$ is one of the states reached during the subset construction applied to the trim part of the reverse of $\mathcal{A}$. We now consider the sets $U_i = \delta^{-1}(U, \alpha^i)$, which are also reached during this subset construction.

Let $\mathcal{C} = \cup_{j \in [m]} \mathcal{C}_j$ and let $\sigma$ be the permutation of $\mathcal{C}$ defined by $\sigma(x) = y$ if and only if $\delta(y, \alpha) = x$. This permutation is well defined, as every element of $\mathcal{C}$ has a unique preimage by $\alpha$ that lies in $\mathcal{C}$. Observe that $U_i \cap \mathcal{C} = \sigma^i(U \cap C)$.

We are interested in the natural action of $\langle \sigma \rangle$, the subgroup generated by $\sigma$, on the subsets of $\mathcal{C}$. Each $\mathcal{C}_i$ is stable by $\sigma$ and since it is $u$-primitive, the orbit of $U \cap \mathcal{C}_i$ under the action of $\langle \sigma \rangle$ has size $|\mathcal{C}_i|$. Hence the orbit of $U \cap \mathcal{C}$ has size $\operatorname{lcm}(|\mathcal{C}_1|, \ldots, |\mathcal{C}_m|)$: the sets $U_i \cap \mathcal{C}$'s are pairwise distinct for $0 \leq i \leq \operatorname{lcm}(|\mathcal{C}_1|, \ldots, |\mathcal{C}_m|) - 1$, and so are the sets $U_i$'s. □

Lemma 4 is the first ingredient of our proofs. This is a sufficient condition on the combinatorial structure of an automaton that gives a lower bound on the state complexity of its reverse language.

## 5. Proof of Theorem 2

In this section, we establish our main theorem. From now on, we are working with the ternary alphabet $A = \{a, b, c\}$, as adding more letters just makes the problem easier. The good distributions on $\mathfrak{A}_n$ can be obtained by choosing independently the actions of $a$, $b$ and $c$, and then, independently, the set of final states. By definition,

the actions of the letters are chosen uniformly at random in the set of all possible mappings from $[n]$ to $[n]$. As we shall see in the proof of Theorem 2, the action of $a$ is used to produce large cycles, while the actions of $b$ and $c$ are used to make these cycles primitive and accessible.

### 5.1. *The lcm of the first d cycles of a random permutation*

In this section, we establish that if we take $d$ cycles in a random permutation of $[m]$, the lcm of their length is in $\Theta(m^d)$ with visible probability.

If $\sigma$ is a permutation of $[m]$, its *sequence of cycles* is the ordered sequence of its cycles $(\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_\ell)$, where the cycles are ordered by their smallest element. If $(\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_\ell)$ is the sequence of cycles of $\sigma$ and $d \leq \ell$, the *first d cycles of* $\sigma$ are the cycles $\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_d$. Let $L_d(\sigma) = (|\mathcal{C}_1|, \dots, |\mathcal{C}_d|)$ denote the lengths of the first $d$ cycles of $\sigma$, if $\sigma$ has at least $d$ cycles, and let $L_d(\sigma) = \perp$ otherwise.

**Lemma 5.** *Let $d$ and $m$ be two positive integers. Let $(\ell_1, \dots, \ell_d) \in [\![\frac{m}{3d}, \frac{m}{2d}]\!]^d$. For the uniform distribution on $\mathfrak{S}_m$ and $m$ sufficiently large we have:*

$$\mathbb{P}\big(L_d = (\ell_1, \dots, \ell_d)\big) \geq \frac{1}{m^d}.$$

**Proof.** Recall that the length of the first cycle of a random permutation of $[m]$ is uniformly distributed: for every $j \in [m]$, the probability that it has length $j$ is $\frac{1}{m}$. Conditioned by the size $\ell_1$ of its first cycle, the remainder of the permutation is a uniform permutation on a set of size $m - \ell_1$. Hence, the probability that $\mathcal{C}_2$ has length $\ell_2$ given that $\mathcal{C}_1$ has length $\ell_1$ is $\frac{1}{m-\ell_1}$. Therefore, by direct induction, the following equality holds for $m$ sufficiently large:

$$\mathbb{P}\big(L_d = (\ell_1, \dots, \ell_d)\big) = \frac{1}{m(m-\ell_1)(m-\ell_1-\ell_2)\cdots(m-\ell_1-\cdots-\ell_{d-1})}.$$

This yields the announced lower bound. □

We now focus on the lcm of the lengths of the first $d$ cycles of a random permutation.

**Proposition 6.** *Let $(x_1, \dots, x_d)$ be a uniform element of $[\![\frac{m}{3d}, \frac{m}{2d}]\!]^d$. There exists a constant $\lambda > 0$ such that $\operatorname{lcm}(x_1, \dots, x_d) \geq \lambda m^d$ with visible probability.*

**Proof.** Let $I_m = [\![\frac{m}{3d}, \frac{m}{2d}]\!]$. We are interested in the uniform distribution on $I_m^d$, the set of $d$-tuples of elements of $I_m$.

For any $p \geq 2$ one can prove by direct counting that the probability $\mathbb{P}_m(p \text{ divides})$ that $p$ divides a uniform random element of $I_m$ satisfies

$$\left|\mathbb{P}_m(p \text{ divides}) - \frac{1}{p}\right| \leq \frac{1}{|I_m|}. \tag{1}$$

Let $Z_m$ denote the random variable defined by on $I_m^d$ by

$$Z_m(x_1, \ldots, x_d) = \sum_{p \text{ divides at least one } x_i} \log p,$$

where $p$ range over the prime numbers smaller than or equal to $|I_m|$. This random variable can be decomposed into (dependent) random variables

$$Z_m = X_{2,m} + \ldots + X_{p_m,m},$$

where $p_m$ is the greatest prime number smaller than or equal to $m$ and

$$X_{p,m}(x_1, \ldots, x_d) = \begin{cases} \log p & \text{if } p \text{ is prime and divides at least one } x_i, \\ 0 & \text{otherwise.} \end{cases}$$

By Equation (1), the probability that a prime number $p$ divides at least one $x_i$ satisfies

$$\begin{aligned} \mathbb{P}\left(p \text{ divides at least one } x_i\right) &= 1 - \mathbb{P}\left(p \text{ divide no } x_i\right) \\ &= 1 - \left(1 - \mathbb{P}_m\left(p \text{ divides}\right)\right)^d \\ &\geq 1 - \left(1 - \frac{1}{p} + \frac{1}{|I_m|}\right)^d. \end{aligned}$$

For $x \in [0,1]$, we have $(1-x)^d \leq 1 - dx + \frac{d(d-1)}{2}x^2$, thus

$$\begin{aligned} \mathbb{P}\left(p \text{ divides at least one } x_i\right) &\geq \frac{d}{p} - \frac{d}{|I_m|} - \frac{d(d-1)}{2}\left(\frac{1}{p} - \frac{1}{|I_m|}\right)^2 \\ &\geq \frac{d}{p} - \frac{d}{|I_m|} - \frac{d(d-1)}{2}\frac{1}{p^2}. \end{aligned}$$

Hence, for $p \leq |I_m|$ we have

$$\mathbb{E}\left[X_{p,m}\right] \geq \frac{d\log p}{p} - \frac{d\log p}{|I_m|} + \mathcal{O}\left(\frac{\log p}{p^2}\right).$$

Thus, by linearity of the expectation

$$\mathbb{E}\left[Z_n\right] \geq \sum_{p \text{ prime} \leq |I_m|} \left(\frac{d\log p}{p} - \frac{d\log p}{|I_m|} + \mathcal{O}\left(\frac{\log p}{p^2}\right)\right).$$

The asymptotic of the first term is given by Mertens' first theorem (see Section 1.4 of [15]):

$$\sum_{p \text{ prime} \leq |I_m|} \frac{d\log p}{p} = d\log|I_m| + \mathcal{O}(1).$$

For the second term, we use the fact that the number of primes smaller than or equal to $x$ grows in $\frac{x}{\log x}$, so that it is smaller than $\frac{2x}{\log x}$ for $x$ sufficiently large:

$$\sum_{p \text{ prime} \leq |I_m|} \log p \leq 2\frac{|I_m|}{\log|I_m|}\log|I_m| \leq 2|I_m|,$$

which yields

$$\sum_{p \text{ prime} \leq |I_m|} \frac{d \log p}{|I_m|} \leq 2d = \mathcal{O}(1),$$

The third term is also in $\mathcal{O}(1)$ as the series converges. Hence we have

$$\mathbb{E}[Z_m] \geq d \log |I_m| + O(1),$$

so that for some positive $C$ and $m$ sufficiently large, we have $\exp(\mathbb{E}[Z_m]) \geq C \cdot m^d$.

Since the function $x \mapsto \exp(x)$ is convex, Jensen's inequality applies and we have $\exp(\mathbb{E}[Z_m]) \leq \mathbb{E}[\exp(Z_m)]$. Moreover,

$$\exp(Z_m(x_1, \ldots, x_d)) = \exp\Big(\sum_{\substack{p \text{ divides at} \\ \text{least one } x_i}} \log p\Big) = \prod_{\substack{p \text{ divides at} \\ \text{least one } x_i}} p \leq \operatorname{lcm}(x_1, \ldots, x_d).$$

And therefore, $\mathbb{E}[\operatorname{lcm}(x_1, \ldots, x_d)] \geq C \cdot m^d$.

We can now conclude the proof, writing "lcm" for $\operatorname{lcm}(x_1, \ldots, x_d)$ to simplify the notations:

$$\begin{aligned}\mathbb{E}[\text{lcm}] &\leq \mathbb{P}\left(\text{lcm} \leq \frac{1}{2}Cm^d\right) \cdot \frac{1}{2}Cm^d + \mathbb{P}\left(\text{lcm} > \frac{1}{2}Cm^d\right) \cdot m^d \\ &\leq \frac{1}{2}Cm^d + \mathbb{P}\left(\text{lcm} > \frac{1}{2}Cm^d\right) \cdot m^d.\end{aligned}$$

And thus

$$\mathbb{P}\left(\operatorname{lcm}(x_1, \ldots, x_d) > \frac{1}{2}Cn^d\right) \geq \frac{1}{2}C,$$

yielding the announced statement by setting $\lambda = \frac{1}{2}C$. □

Let $\text{Cycle}_d(m)$ denote the set of permutations $\sigma$ of $\mathfrak{S}_m$ such that $L_d(\sigma) \in [\![\frac{m}{3d}, \frac{m}{2d}]\!]^d$ and $\operatorname{lcm}(\ell_1, \ldots, \ell_d) \geq \lambda m^d$, with $L_d(\sigma) = (\ell_1, \ldots, \ell_d)$. We use the $\lambda$ of Proposition 6.

If $\sigma$ is a random permutation conditioned by $L_d(\sigma) \in [\![\frac{n}{3d}, \frac{n}{2d}]\!]^d$, the vector $L_d(\sigma)$ is not uniformly distributed in $[\![\frac{m}{3d}, \frac{m}{2d}]\!]^d$. However, we can control the lack of uniformity and use Proposition 6 to obtain sufficiently many permutations such that the lcm of their first $d$ cycles is sufficiently large.

**Lemma 7.** *For any positive integer $d$, a uniform random permutation of $[m]$ is in $\text{Cycle}_d(m)$ with visible probability.*

**Proof.** Let $E_d(m)$ denote the set of elements $(\ell_1, \ldots, \ell_d) \in [\![\frac{m}{3d}, \frac{m}{2d}]\!]^d$ such that we have $\operatorname{lcm}(\ell_1, \ldots, \ell_d) \geq \lambda m^d$. By Proposition 6, there exists $C > 0$ such that $|E_d(m)| \geq C\big|[\![\frac{m}{3d}, \frac{m}{2d}]\!]^d\big|$ for $m$ sufficiently large. Hence, by Lemma 5,

$$\mathbb{P}(\text{Cycle}_d(m)) = \sum_{(\ell_1, \ldots, \ell_d) \in E_d(m)} \mathbb{P}\big(L_d = (\ell_1, \ldots, \ell_d)\big) \geq \frac{|E_d(m)|}{m^d} \geq C \frac{\big|[\![\frac{m}{3d}, \frac{m}{2d}]\!]\big|^d}{m^d}.$$

This conclude the proof, as $\frac{1}{m}\big|[\![\frac{m}{3d}, \frac{m}{2d}]\!]\big|$ tends to $\frac{1}{6d}$. □

### 5.2. *Some properties of random mappings*

In this section we establish two different properties of random mappings. By first working on the cyclic part of the random mapping associated with the action of $a$, we exhibit some $a$-cycles such that the lcm of their lengths is sufficiently large (using the results of Section 5.1). Next, we focus on the largest tree in the decomposition of the action of $b$; this will be needed to prove that the $a$-cycles we obtained are primitive sufficiently often, which is required in order to apply Lemma 4.

#### 5.2.1. *Some properties of the cyclic part of a random mapping*

We first use classical techniques of analytic combinatorics to obtain some information on the size of the cyclic part of a random mapping.

**Lemma 8.** *The cyclic part of a uniform random mapping of size $n$ has size in $[\![\sqrt{n}, 2\sqrt{n}]\!]$ with visible probability.*

**Proof.** Let $C_n$ denote the random variable that counts the number of cyclic points in a mapping of size $n$. Classically, the exponential bivariate generating function of mappings where cycles are marked is (see [11]):

$$M(z,u) := \sum_{n,k} \frac{m_{n,k}}{n!} z^n u^k = \frac{1}{1-uT(z)}, \quad \text{with } T(z) = z\exp(T(z)),$$

where $m_{n,k}$ is the number of size $n$ mappings with $k$ cycles and $T(z)$ is the exponential generating function of (rooted) Cayley trees. From [10], the unique dominant singularity of $T(z)$ is $\rho := e^{-1}$ and near $\rho$ the following development holds:

$$T(z) = 1 - \sqrt{2}\sqrt{1-ez} + o\left(\sqrt{1-ez}\right).$$

To obtain the expected number of cycles, we compute the derivative of $M(z,u)$ with respect to $u$:

$$\frac{d}{du}M(z,u)\Big|_{u=1} = \frac{T(z)}{(1-T(z))^2} = \frac{1}{2(1-ez)} + o\left(\frac{1}{1-ez}\right).$$

If $[z^n]A(z)$ denote the $n$-th coefficient of the series $A(z)$, the Transfer Theorem (Corollary VI.1 p. 392 of [11]) yields, using the notation $M(z) = M(z,1)$,

$$[z^n]\frac{d}{du}M(z,u)\Big|_{u=1} \sim \frac{e^n}{2} \quad \text{and} \quad [z^n]M(z) \sim \frac{e^n}{\sqrt{2\pi n}}.$$

Thus, the expected number of cyclic points satisfies:

$$\mathbb{E}[C_n] = \frac{[z^n]\frac{d}{du}M(z,u)\Big|_{u=1}}{[z^n]M(z)} \sim \sqrt{\frac{\pi n}{2}}.$$

We now compute the variance of the number of cyclic points using the following formula for the second moment of $C_n$:

$$\mathbb{E}[C_n^2] = \frac{[z^n]\frac{d^2}{du^2}M(z,u)\Big|_{u=1} + [z^n]\frac{d}{du}M(z,u)\Big|_{u=1}}{[z^n]M(z)}.$$

We have

$$\frac{d^2}{du^2}M(z,u)\Big|_{u=1} = \frac{2T(z)^2}{(1-T(z))^3} = \frac{1}{\sqrt{2}(1-ez)^{3/2}} + o\left(\frac{1}{(1-ez)^{3/2}}\right).$$

Hence, using the Transfer Theorem again:

$$[z^n]\frac{d^2}{du^2}M(z,u)\Big|_{u=1} \sim \sqrt{\frac{2n}{\pi}}e^n \quad \text{and} \quad \mathbb{E}[C_n^2] \sim 2n.$$

This yields the following asymptotic equivalent for the variance of $C_n$:

$$\mathbb{V}[\mathcal{C}_n] = \mathbb{E}[C_n^2] - \mathbb{E}[C_n]^2 \sim \frac{4-\pi}{2}\, n.$$

The result follows by Chebyshev's inequality,[f] as $\sqrt{n} \leq \sqrt{\pi\, n/2} \leq 2\sqrt{n}$. □

Let $\mathcal{G}(n,d)$ denote the set of mappings $f$ of size $n$ such that:

(1) The number $m$ of cyclic points of $f$ is in $[\![\sqrt{n}, 2\sqrt{n}]\!]$.
(2) The normalized cyclic permutation of $f$ is in $\mathrm{Cycle}_d(m)$.

**Proposition 9.** *A mapping of size $n$ is in $\mathcal{G}(n,d)$ with visible probability. Moreover, for $n$ sufficiently large, the first $d$ cycles of every element of $\mathcal{G}(n,d)$ have sizes in $[\![\frac{\sqrt{n}}{3}, \sqrt{n}]\!]$ and the lcm of their lengths is at least $\lambda n^{d/2}$.*

**Proof.** By definition, if $\sigma$ denote the normalized cyclic permutation we have

$$\begin{aligned}\mathbb{P}\left(\mathcal{G}(n,d)\right) &= \sum_{m\in[\![\sqrt{n},2\sqrt{n}]\!]} \mathbb{P}\left(C_n = m,\ \sigma \in \mathrm{Cycle}_d(m)\right) \\ &= \sum_{m\in[\![\sqrt{n},2\sqrt{n}]\!]} \mathbb{P}\left(\sigma \in \mathrm{Cycle}_d(m) \mid C_n = m\right) \cdot \mathbb{P}\left(C_n = m\right).\end{aligned}$$

As conditioned by its size, the normalized cyclic permutation of a uniform random mapping is a uniform permutation, Lemma 7 directly yields that for $m$ sufficiently large, $\mathbb{P}\left(\sigma \in \mathrm{Cycle}_d(m) \mid C_n = m\right) \geq \delta$, for some positive $\delta$. Thus,

$$\mathbb{P}\left(\mathcal{G}(n,d)\right) \geq \delta \sum_{m\in[\![\sqrt{n},2\sqrt{n}]\!]} \mathbb{P}\left(C_n = m\right) = \delta\ \mathbb{P}\left(C_n \in [\![\sqrt{n}, 2\sqrt{n}]\!]\right).$$

By Lemma 8, a mapping is therefore in $\mathcal{G}(n,d)$ with visible probability. The last part of the statement follows from the definitions of $\mathcal{G}(n,d)$ and $\mathrm{Cycle}_d(m)$. □

[f] Note that the distribution is not concentrated around is mean, since the expectation and standard deviation have the same order of growth in $\sqrt{n}$.

5.2.2. *Some properties of the largest tree of a random mapping*

Recall that a rooted Cayley tree is a rooted, non-plane, labelled tree. It means that if there are $n$ nodes, they are labelled with pairwise distinct elements of $[n]$, and that the children of a node are unordered. As mentioned before, a mapping of size $n$ can be seen as a union of cycles of Cayley trees (the roots of the trees are cyclically linked and form the cyclic permutation of the mapping).

Define the *largest tree* of a mapping $f$ as the largest Cayley tree of its decomposition, taking the tree with the smallest root label if there are several trees with the maximum number of nodes. Our next lemma states that conditioned by its size, the largest tree of a random mapping behaves like a uniform random tree.

**Lemma 10.** *Let $t$ and $n$ be two integers such that $1 \leq \frac{n}{2} < t \leq n$ and let $\mathfrak{M}_n^{(t)}$ denote the set of mapping on $[n]$ whose largest tree has $t$ nodes. The normalized largest tree of a uniform element of $\mathfrak{M}_n^{(t)}$ is distributed as a uniform random Cayley tree with $t$ nodes.*

**Proof.** Let $T$ and $T'$ be two Cayley trees with $t$ nodes. Let $\phi$ denote the map from $\mathfrak{M}_n^{(t)}$ into itself that operates as follows: if the largest tree normalizes in $T$, we change the shape of the tree so that it now normalizes in $T'$ (there is a unique way to do that), and conversely if it normalizes in $T'$ we change it so that it normalizes to $T$ (there is also a unique way to do that). If the largest tree does not normalize in $T$ or $T'$, the mapping remains unchanged by $\phi$. As an involution, the map $\phi$ is one-to-one. Hence, when $f$ is a uniform random mapping, so is $\phi(f)$. The probability that the largest tree normalizes in $T$ is the therefore the same that it normalizes in $T'$, concluding the proof. □

For any $w > 0$, let $\mathcal{H}(n, d, w)$ denote the set of mappings of size $n$ such that the largest Cayley tree of its decomposition is of width at least $w\sqrt{n}$.

Proposition 11 below is established using a classical result on the size of the largest tree of a random mappings [13] and the analysis of the width of a random Cayley tree [5].

**Proposition 11.** *There exists a positive real number $w$ such that a mapping of size $n$ is in $\mathcal{H}(n, d, w)$ with visible probability.*

**Proof.** From [13] p. 164, we get that with positive probability, the largest tree of a random mapping has more than $\frac{2}{3}n$ nodes. By Lemma 10, conditioned by its size $t$, this largest tree behaves like a uniform random Cayley tree with $t$ nodes. By [5], the width of such a Cayley tree is greater than $w'\sqrt{t}$ with positive probability, for some $w' > 0$. Taking $w = w'\sqrt{2/3}$ yields that the width of the largest tree is at least $w\sqrt{n}$ with visible probability, concluding the proof. □

### 5.3. *Some properties of random transition structures*

In this section, we use well established results on random deterministic automata [14] to obtain a word $v$ on $A' = \{b, c\}$ such that $\delta^{-1}(f, v)$ is sufficiently large, with visible probability, where $f$ is a random final state.

A set of vertices $X$ of a digraph is *stable* if there is no edge $x \to y$ for $x \in X$ and $y \notin X$. We shall need the following results in the sequel:

**Lemma 12.** *Let $\mathcal{A}$ be a transition structure of $\mathfrak{T}_n(A')$ taken uniformly at random. Let $G$ be the underlying digraph induced on $[n]$ by the actions of $b$ and $c$ (there is an edge $x \to y$ if and only if $\delta(x, b) = y$ or $\delta(x, c) = y$). With high probability, $G$ has a unique stable strongly connected component, which has size greater than $\frac{1}{2}n$.*

**Proof.** We first prove that, with high probability, there is no stable set of states of size smaller than $\frac{1}{4}n$: we overcount the number of transition structures having a stable subset $X$ of size $\ell$ by choosing the $\ell$ states, their images by both letters in $X$ and the images of the other states. This yields an upper bound of $\binom{n}{\ell}\ell^{2\ell}n^{2n-2\ell}$ for the number of such transition structures. Hence the probability that there is such a small stable subset is bounded above by:

$$\frac{1}{n^{2n}}\sum_{\ell=1}^{n/4}\binom{n}{\ell}\ell^{2\ell}n^{2n-2\ell} = \sum_{\ell=1}^{n/4}\binom{n}{\ell}\left(\frac{\ell}{n}\right)^{2\ell} \leq \sum_{\ell=1}^{n/4}\left(\frac{en}{\ell}\right)^{\ell}\left(\frac{\ell}{n}\right)^{2\ell} = \sum_{\ell=1}^{n/4}\left(\frac{e\ell}{n}\right)^{\ell},$$

which is $\mathcal{O}(\frac{1}{n})$, as one can see by isolating the term $\ell = 1$.

We now consider the uniform distribution on $\mathfrak{T}_n \times [n]$, where a pair $(\mathcal{A}, i)$ is seen as a transition structure of initial state $i$. In [4] it is proven that in a random transition structure with $n$ states on a two-letter alphabet, the accessible part has size greater than $\frac{1}{2}n$ with high probability. The proof is established when 1 is the initial state, but by symmetry it still holds if we choose the initial state uniformly at random. Let $E_n$ denote the set of transition structures with $n$ states that have a stable strongly connected component of size between $\frac{1}{4}n$ and $\frac{1}{2}n$. If the initial state we add to a transition structure of $E_n$ is inside a stable component of size at most $\frac{1}{2}n$, then the generic property of [4] does not hold. The number of couples $\mathfrak{T}_n \times [n]$ such that the property of [4] does not hold is therefore at least $\frac{n}{4}|E_n|$, and it is also in $o(n|\mathfrak{T}_n|)$. Hence $|E_n| = o(|\mathfrak{T}_n|)$, and generically there is no stable strongly connected component of size smaller than $\frac{1}{2}n$. This conclude the proof, as there can be at most one strongly connected component of size greater $\frac{1}{2}n$. □

We shall also need the following result of [8]:

**Lemma 13 (Proposition 1 in [8])** *Let $A$ be an alphabet with at least 2 letters and let $\alpha \in A$. With high probability, all the $\alpha$-cycles of length greater than $\log n$ are accessible in a uniform random transition structure on $A$.*

The main result of this section is the following proposition.

**Proposition 14.** *Let $\mathcal{A}$ be a uniform random transition structure of $\mathfrak{T}_n(A')$. Let $f$ be a state of $\mathcal{A}$ chosen uniformly in $[n]$ and independently from $\mathcal{A}$. There exists a positive real number $w$ such that, with visible probability, $f$ is accessible and there exists a word $v \in \{b, c\}^*$ such that $\delta^{-1}(f, v)$ has size at least $w\sqrt{n}$.*

**Proof.** We choose $w$ as in Proposition 11. Thus, with visible probability the action of $b$ is in $\mathcal{H}(n, d, w)$ and therefore it contains a tree of width at least $w\sqrt{n}$. Let $r$ be the root of this tree. By Lemma 12, with high probability $\mathcal{A}$ has only one stable strongly connected component $\mathcal{C}$, of size at least $\frac{1}{2}n$. As the intersection of a visible property and a property that holds with high probability is visible, $\mathcal{A}$ satisfies both properties with visible probability.

When a final state is chosen randomly and independently, it is in $\mathcal{C}$ with probability at least $\frac{1}{2}$. Moreover, if the final state is in the unique stable strongly connected component, then there exists a word $u$ that labels a path from $r$ to $f$: when there is a unique stable strongly connected component, it is necessarily accessible from anywhere. Consider the word $v = \tilde{u}b^i$, where $i$ is the layer of $T$ with the maximal number of nodes (the level that gives its width). Then $\delta^{-1}(f, v)$ contains all the states of the $i$-th layer of $T$, and it therefore contains at least $w\sqrt{n}$ elements. By independence of the choice of $f$, this happens with visible probability. □

### 5.4. *Proof of the main theorem*

At this point we have, with visible probability, sufficiently many $a$-cycles to obtain a super-polynomial lower bound, provided we can apply Lemma 4. Lemma 13 ensures that these $a$-cycles are accessible with high probability, so we only have to focus on their primitivity. For this, we will use Proposition 14 and find a word $u$ such that the $a$-cycles are $u$-primitive with visible probability.

The next technical lemma ensures that, with visible probability, $\delta^{-1}(f, v)$ intersects each of the first $d$ $a$-cycles, for the $v$ of Proposition 14:

**Lemma 15.** *Let $x$ and $y$ be two positive real numbers. Let $X$ be a subset of $[n]$ of size $\lceil x\sqrt{n}\,\rceil$ and let $Y$ be a uniform random subset of $[n]$ of size $\lceil y\sqrt{n}\,\rceil$. For every integer $j \geq 0$, there exists a positive constant $M_j$ such that $|X \cap Y| = j$ with probability at least $M_j$, for $n$ sufficiently large.*

**Proof.** Let $m = \lceil x\sqrt{n}\,\rceil$ and $m' = \lceil y\sqrt{n}\,\rceil$. We prove the result by induction on $j$. For $j = 0$, the probability that $X \cap Y = \emptyset$ is

$$\frac{\binom{n-m}{m'}}{\binom{n}{m'}} = \frac{(n-m)!}{n!}\frac{(n-m')!}{(n-m-m')!} \geq \frac{(n-m-m')^m}{n^m} = \left(1 - \frac{m+m'}{n}\right)^m.$$

The quantity on the right tends to $\exp(-x(x+y))$ as $n$ tends to infinity, and is therefore greater than $M_0 = \frac{1}{2}\exp(-x(x+y))$ for $n$ sufficiently large.

By first choosing the elements of the intersection, then the other elements of $Y$, the probability that $|X \cap Y| = j$ is exactly

$$P_j = \frac{\binom{m}{j}\binom{n-m}{m'-j}}{\binom{n}{m'}}.$$

Observe that

$$\frac{P_{j+1}}{P_j} = \frac{\binom{m}{j+1}\binom{n-m}{m'-j-1}}{\binom{m}{j}\binom{n-m}{m'-j}} = \frac{m-j}{j+1} \cdot \frac{m'-j}{n-m-m'+j+1} \xrightarrow[n\to\infty]{} \frac{xy}{j+1}.$$

Hence, if the property holds for $j$, then $P_{j+1} \geq \frac{xy}{2(j+1)} M_j$ for $n$ sufficiently large, concluding the proof. □

Our next lemma ensures that these intersections are not too large, i.e., that we do not have one of the first $d$ $a$-cycles completely included in $\delta^{-1}(f, v)$:

**Lemma 16.** *Let $x$ be a positive real number. Let $X$ be a subset of $[n]$ of size $m = \lceil x\sqrt{n} \rceil$ and let $Y$ be a uniform random subset of $[n]$ of size $m'$ with $1 \leq m' \leq \frac{n}{2}$. For $n$ sufficiently large, the probability that $X \subseteq Y$ is smaller than $n2^{-x\sqrt{n}}$.*

**Proof.** We add some random elements to $Y$ until we have a set $Y'$ of size $\ell = \lfloor \frac{n}{2} \rfloor$. The probability that $Y$ contains $X$ is smaller than the probability that $Y'$ contains $X$, which is exactly

$$\frac{\binom{n-m'}{\ell-m}}{\binom{n}{\ell}} \leq \frac{2^{n-m}}{\binom{n}{\ell}},$$

which concludes the proof since $\binom{n}{\ell} \sim \sqrt{\frac{2}{\pi n}} 2^n$. □

Our last lemma will be used to prove that when $\delta^{-1}(f, v)$ intersects non-trivially the first $d$ $a$-cycles, then they are primitive with high probability.

**Lemma 17.** *Let $n \geq 2$. Let $\mu$ be a probability mass on $\{0,1\}^n$ such that $\mu(0^n) = \mu(1^n) = 0$ and such that two words with the same number of 0's have the same probability. Then the probability that a word is not primitive for $\mu$ is at most $\frac{2}{n}$.*

**Proof.** Let $L_n$ be the random variable that count the number of 0's. By symmetry between 0's and 1's in the definition of primitivity, we work for $L_n = i$ with $1 \leq i \leq \frac{n}{2}$ only. The conditions on $\mu$ imply that, conditioned by $L_n = i$, we have the uniform distribution on the words with $i$ occurrences of 0.

If $u$ is not primitive, then $u = v^\ell$ for some $\ell \in [\![2, i]\!]$ that divides both $i$ and $n$. For every such $\ell$ and every $v$ of length $n/\ell$ with $i/\ell$ occurrences of 0, there is exactly one non-primitive $u$ of the form $v^\ell$, but there are $\binom{n-n/\ell}{i-i/\ell}$ words with the correct number of 0's whose prefix is $v$. Hence for any valid $\ell$, the probability that $u = v^\ell$ is bounded from above by $\binom{n-n/\ell}{i-i/\ell}^{-1}$. When $i = 2$, $\ell = 2$ is the only possible

value for $\ell$, yielding an upper bound of $\binom{n-n/2}{1}^{-1} = \frac{2}{n}$ for even $n$. If $i \geq 3$, then $i - i/\ell \geq 2$ for valid values of $\ell$ and $n - n/\ell \geq \frac{n}{2}$. Thus, using the properties of the binomial coefficients, for each valid value of $\ell$ the upper bound is at most $\frac{8}{n(n-2)}$. This concludes the proof for any given value of $i$, as there are at most $n-2$ valid values for $\ell$. Since the bound does not depend on $i$, this yields the result. □

We can now prove our main result.

**Proof of Theorem 2.** By definition of a good distribution with a small number of final states, if $F_n$ is the random variable that counts the number of final states, then $F_n \in [\![1, \frac{n}{2}]\!]$ with visible probability. We first condition on $F_n = n_f$, for any $n_f \in [\![1, \frac{n}{2}]\!]$ such that $\mathbb{P}(F_n = n_f) \neq 0$. As final states and transition structures are chosen independently, we are interested in the uniform distribution on transition structures on $A = \{a, b, c\}$ having exactly $n_f$ final states. We also denote by $f$ a final state chosen uniformly at random amongst the set of final states. By definition of a label-independent distribution, $f$ follows a uniform distribution on $[n]$.

By Proposition 9, the action of $a$ is in $\mathcal{G}(n, d)$ with probability at least $C_a$, for some positive $C_a$, when $n$ is sufficiently large. By Proposition 14, $f$ is accessible in the automaton restricted to the transitions labelled by $b$'s and $c$'s, and there exists a word $v \in \{b, c\}^*$ such that $\delta^{-1}(f, v)$ has size at least $w\sqrt{n}$ with probability at least $C_{bc}$, for some positive $C_{bc}$, when $n$ is sufficiently large. By independence of the action of the letters in a uniform random element of $\mathfrak{T}_n$, and by independence of $f$, all these properties hold with probability at least $C_a C_{bc}$ for $n$ sufficiently large. Let $\mathcal{X}_n$ denote the set of automata that satisfy these conditions.

We now consider the uniform distribution on elements of $\mathcal{X}_n$ with $n_f$ final states. By definition of $\mathcal{G}(n, d)$, the first $d$ cycles $\mathcal{C}_1, \ldots, \mathcal{C}_d$ of $\delta_a$ have sizes at least $\lceil \frac{\sqrt{n}}{3} \rceil$. Let $S_i$ be a uniform subset of size $\lceil \frac{\sqrt{n}}{3} \rceil$ of each such $\mathcal{C}_i$. Observe that if a map $\phi \in \mathcal{G}(n, d)$, then every relabelling of $\phi$, i.e. considering $\sigma \circ \phi \circ \sigma^{-1}$ for a permutation $\sigma \in \mathfrak{S}_n$, is also in $\mathcal{G}(n, d)$. Hence, $X := \cup_{i=1}^{d} S_i$ is a uniform random subset of $[n]$ of size $d\lceil \frac{\sqrt{n}}{3} \rceil$.

Let $F'$ be the set of accessible final states. As $\delta^{-1}(f, v)$ has size at least $w\sqrt{n}$ and $f \in F'$, $\delta^{-1}(F', v)$ also has size at least $w\sqrt{n}$, as $f$ is accessible. Moreover, $\delta^{-1}(F', \varepsilon)$ has size at most $n_f \leq \frac{n}{2}$. Therefore, there exists a prefix $u$ of $v$ such that $\delta^{-1}(F', u)$ has size in $[\![w\sqrt{n}, \frac{n}{2}]\!]$.

Let $Y$ be a uniform random subset of size $\lceil w\sqrt{n} \rceil$ of $\delta^{-1}(F', u)$. By Lemma 15, $|X \cap Y| = d$ with visible probability. A direct computation shows that in this case, $Y$ intersects each $S_i$ exactly once also with visible probability. Moreover, Lemma 16 ensures that no $S_i$ is included in $\delta^{-1}(F', \varepsilon)$ with high probability.

As a conclusion, for the uniform distribution on automata of $\mathcal{X}_n$ with $n_f$ final states, $\delta^{-1}(F', \varepsilon)$ intersects non-trivially each $\mathcal{C}_i$, with visible probability. By independence of the actions of the letters in a uniform element of $\mathcal{X}_n$, we can therefore apply Lemma 17, and obtain that with probability at least $C$, every $\mathcal{C}_i$ is primi-

tive, for some positive constant $C$ and $n$ sufficiently large. Importantly, $C$ does not depend on $n_f$, as everything is done starting from one final state $f$.

We conclude the proof as follows. For $n$ sufficiently large, the $\mathcal{C}_i$'s are all primitive for a constant proportion of automata in $\mathcal{X}_n$ with $n_f$ final states. But this set of automaton is also a constant proportion of the automata with $n_f$ final states. Hence, a uniform random automaton with $n_f$ states has this property with probability at least $C_a C_{bc} C$, which does not depend on $n_f$. Let $\mathcal{Y}_n$ denote the subset of automata in $\mathcal{X}_n$ such that the $\mathcal{C}_i$'s are all primitive. By the law of total probabilities, we have

$$\begin{aligned}\mathbb{P}(\mathcal{A} \in \mathcal{Y}_n) &= \sum_{n_f \in [\![0,n]\!]} \mathbb{P}(\mathcal{A} \in \mathcal{Y}_n \mid F_n = n_f) \cdot \mathbb{P}(F_n = n_f) \\ &\geq \sum_{n_f \in [\![1,n/2]\!]} \mathbb{P}(\mathcal{A} \in \mathcal{Y}_n \mid F_n = n_f) \cdot \mathbb{P}(F_n = n_f) \\ &\geq C_a C_{bc} C \sum_{n_f \in [\![1,n/2]\!]} \mathbb{P}(F_n = n_f).\end{aligned}$$

The last sum is exactly the probability that there are between 1 and $\frac{n}{2}$ final states. By definition of a distribution with a small number of final states, this is visible, and thus an automaton is in $\mathcal{Y}_n$ with visible probability.

By Lemma 13, the $\mathcal{C}_i$'s are also accessible with high probability. Hence, even if we only keep the trim part of the random automaton, there is a path from the initial state to the $\mathcal{C}_i$'s, which are $u$-primitive. By Lemma 4, this yields a lower bound of $\lambda n^{d/2}$ for the state complexity of the reverse. This concludes the proof by taking $d \geq 2\gamma + 1$. □

## 6. Conclusions

A natural question is whether the average super-polynomial running time of Brzozowski's algorithm still holds for alphabets with two letters, for good distributions of automata with a small number of final states. The proof of this paper relies on the fact that we built $u \in \{b, c\}^*$ and the $a$-cycles independently. This proof cannot be adapted to binary alphabet without taking into account the dependencies between a word $u$ using the letter $a$ and the $a$-cycles. This is probably quite difficult to handle. A completely different approach is probably required in order to obtain the generalization to binary alphabets.

Another interesting direction would be to give some insight on distributions on accessible (or even trim) automata, instead of on any deterministic and complete automata. This is possibly much more difficult, as the classical tool to prove such results fails here (see Section 3 of [14]).

**Acknowledgment.** We would like to thanks Jean-François Marckert for his crystal clear explanation of his result on the width of random Cayley trees [5]. We also thank the referees for providing helpful comments, which helped to improve the quality of this article.

## References

[1] F. Bassino, J. David, and C. Nicaud. Average case analysis of Moore's state minimization algorithm. *Algorithmica*, 63(1-2):509–531, 2012.

[2] F. Bassino, J. David, and A. Sportiello. Asymptotic enumeration of minimal automata. In Dürr and Wilke [7], pages 88–99.

[3] J. A. Brzozowski. Canonical regular expressions and minimal state graphs for definite events. In *Mathematical theory of Automata*, pages 529–561. Polytechnic Press, Polytechnic Institute of Brooklyn, N.Y., 1962. Volume 12 of MRI Symposia Series.

[4] A. Carayol and C. Nicaud. Distribution of the number of accessible states in a random deterministic automaton. In Dürr and Wilke [7], pages 194–205.

[5] P. Chassaing and J.-F. Marckert. Parking functions, empirical processes, and the width of rooted labeled trees. *Electron. J. Combin*, 8(1), 2001.

[6] J. David. Average complexity of Moore's and Hopcroft's algorithms. *Theor. Comput. Sci.*, 417:50–65, 2012.

[7] C. Dürr and T. Wilke, editors. *29th International Symposium on Theoretical Aspects of Computer Science, STACS 2012, February 29th - March 3rd, 2012, Paris, France*, volume 14 of *LIPIcs*. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2012.

[8] S. D. Felice and C. Nicaud. Brzozowski algorithm is generically super-polynomial for deterministic automata. In M.-P. Béal and O. Carton, editors, *Developments in Language Theory*, volume 7907 of *Lecture Notes in Computer Science*, pages 179–190. Springer, 2013.

[9] S. D. Felice and C. Nicaud. On the average complexity of Brzozowski's algorithm for deterministic automata with a small number of final states. In A. M. Shur and M. V. Volkov, editors, *Developments in Language Theory - 18th International Conference, DLT 2014, Ekaterinburg, Russia, August 26-29, 2014. Proceedings*, volume 8633 of *Lecture Notes in Computer Science*, pages 25–36. Springer, 2014.

[10] P. Flajolet and A. M. Odlyzko. Random mapping statistics. In J.-J. Quisquater and J. Vandewalle, editors, *EUROCRYPT*, volume 434 of *Lecture Notes in Computer Science*, pages 329–354. Springer, 1989.

[11] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.

[12] J. E. Hopcroft. An $n \log n$ algorithm for minimizing the states in a finite automaton. In Z. Kohavi, editor, *The Theory of Machines and Computations*, pages 189–196. Academic Press, 1971.

[13] V. Kolčin. *Random Mappings: Translation Series in Mathematics and Engineering*. Translations series in mathematics and engineering. Springer London, Limited, 1986.

[14] C. Nicaud. Random deterministic automata. In E. Csuhaj-Varjú, M. Dietzfelbinger, and Z. Ésik, editors, *Mathematical Foundations of Computer Science 2014 - 39th International Symposium, MFCS 2014, Budapest, Hungary, August 25-29, 2014. Proceedings, Part I*, volume 8634 of *Lecture Notes in Computer Science*, pages 5–23. Springer, 2014.

[15] G. Tenenbaum. *Introduction to Analytic and Probabilistic Number Theory*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1995.

**Appendix: answer to the referees' comments**

We first want to thanks the referees and the editor for their very useful comments. The article is much better now.

We took all the comments into account in the current version of the paper, and reorganized the subsections of Section 5 as suggested. We just chose to keep the notion of "state complexity", which is quite standard in automata theory: we added its definition at the beginning of the article, and reworded some statements to make it appear everywhere it was meaningful.