Minimal automaton for multiplying and translating the Thue-Morse set

Émilie Charlier   Célia Cisternino *   Adeline Massuir

Department of Mathematics
University of Liège
Belgium

`{echarlier,ccisternino,a.massuir}@uliege.be`

Abstract

The Thue-Morse set $\mathcal{T}$ is the set of those non-negative integers whose binary expansions have an even number of 1's. The name of this set comes from the fact that its characteristic sequence is given by the famous Thue-Morse word

$$0110100110010110\cdots,$$

which is the fixed point starting with 0 of the word morphism $0 \mapsto 01, 1 \mapsto 10$. The numbers in $\mathcal{T}$ are commonly called the *evil numbers*. We obtain an exact formula for the state complexity of the set $m\mathcal{T} + r$ (i.e. the number of states of its minimal automaton) with respect to any base b which is a power of 2. Our proof is constructive and we are able to explicitly provide the minimal automaton of the language of all 2^p-expansions of the set of integers $m\mathcal{T} + r$ for any positive integers p and m and any remainder $r \in \{0, \ldots, m-1\}$. The proposed method is general for any b-recognizable set of integers.

Mathematics Subject Classifications: 68Q70,11A67,11B85

1 Introduction

A subset X of $\mathbb{N}$ is said to be *b-recognizable* if the base-b expansions of the elements of X form a regular language. The famous theorem of Cobham tells us that any non-trivial recognizable property of numbers is dependent on the base we choose: the only sets that are b-recognizable for all bases b are the finite unions of arithmetic progressions [19].

Inspired by this seminal result, many descriptions of b-recognizable sets were given, e.g. morphic, algebraic and logical characterizations [9, 11, 20], extensions of these to systems based on a Pisot number [10], the normalization map [23] or the possible growth functions [16, 22]. For more on b-recognizable sets, we refer to the surveys [7, 11, 13, 22, 24, 35].

In particular, as mentioned above, these sets have been characterized in terms of logic. More precisely, a subset of $\mathbb{N}$ (and more generally of $\mathbb{N}^d$) is b-recognizable if and only if it is definable by a first-order formula of the structure $\langle \mathbb{N}, +, V_b \rangle$ where V_b is the base-dependent functional predicate that associates with a natural n the highest power of b dividing n. Since the finite unions of arithmetic progressions are precisely the subsets of $\mathbb{N}$ that are definable by first order formulas in the Presburger arithmetic $\langle \mathbb{N}, + \rangle$, this characterization provides us with a logical interpretation of Cobham's theorem. In addition, this result turned out to be a powerful tool for showing that many properties of b-automatic sequences are decidable and, further, that many enumeration problems of b-automatic sequences can be described by b-regular sequences in the sense of Allouche and Shallit [5, 7, 18].

In the context of Cobham's theorem, the following question is natural and has received constant attention during the last 30 years: given an automaton accepting the language of the base-b expansions of a set $X \subseteq \mathbb{N}$, is it decidable whether X is a finite union of arithmetic progressions? Several authors gave decision procedures for this problem [4, 11, 25, 26, 28]. Moreover, a multidimensional version of this problem was shown to be decidable in a beautiful way based on logical methods [11, 33].

With any set of integers X is naturally associated an infinite word, which is its characteristic sequence $\chi_X \colon n \mapsto 1$ if $n \in X$, $n \mapsto 0$ otherwise. Thus, to a finite union of arithmetic progressions corresponds an ultimately periodic infinite word. Therefore, the HD0L ultimate periodicity problem consisting in deciding whether a given morphic word (i.e. the image under a coding of the fixed point of a morphism) is ultimately periodic is a generalization of the periodicity problem for b-recognizable sets mentioned in the previous paragraph. The HD0L ultimate periodicity problem was shown to be decidable in its full generality [21, 31]. The proofs rely on return words, primitive substitutions or evolution of Rauzy graphs. However, these methods do not provide algorithms that could be easily implemented and the corresponding time complexity is very high. In addition, they do not allow us to obtain an algorithm for the multidimensional generalization of the periodicity problem, i.e. the problem of deciding whether a b-recognizable subset of $\mathbb{N}^d$ is definable within the Presburger arithmetic $\langle \mathbb{N}, + \rangle$. Therefore, a better understanding of the inner structure of automata arising from number systems remains a powerful tool to obtain efficient decision procedures.

The general idea is as follows. Suppose that $\mathcal{L} = \{L_i \colon i \in \mathbb{N}\}$ is a collection of languages and that we want to decide whether some particular language L belongs to $\mathcal{L}$. Now, suppose that we are able to explicitly give a lower bound on the state complexities of the languages in $\mathcal{L}$, i.e. for each given N, we can effectively produce a bound $B(N)$ such that for all $i > B(N)$, the state complexity of L_i is greater than N. Then the announced problem is decidable: if N is the state complexity of the given language L, then only the finitely many languages $L_0, \ldots, L_{B(N)}$ have to be compared with L.

The state complexity of a b-recognizable set (i.e. the number of states of the minimal

automaton accepting the b-expansions of its elements) is closely related to the length of the logical formula describing this set. Short formulas are crucial in order to produce efficient mechanical proofs by using for example the Walnut software [32, 37]. There are several ways to improve the previous decision procedure. One of them is to use precise knowledge of the structure of the involved automata. This idea was successfully used in the papers [8, 28]. In [17], the structure of automata accepting the greedy expansions of $m\mathbb{N}$ for a wide class of non-standard numeration systems, and in particular, estimations of the state complexity of $m\mathbb{N}$ are given. Another way of improving this procedure is to have at our disposal the exact state complexities of the languages in $\mathcal{L}$. Finding an exact formula is a much more difficult problem than finding good estimates. However, some results in this direction are known. For instance, it is proved in [17] that for the Zeckendorf numeration system (i.e. based on the Fibonacci numbers), the state complexity of $m\mathbb{N}$ is exactly $2m^2$. A complete description of the minimal automaton recognizing $m\mathbb{N}$ in any integer base b was given in [1] and the state complexity of $m\mathbb{N}$ with respect to the base b is shown to be exactly

$$\frac{m}{\gcd(m, b^N)} + \sum_{t=0}^{N-1} \frac{b^t}{\gcd(m, b^t)}$$

where N is the smallest integer α such that $\frac{m-b^\alpha}{\gcd(m,b^\alpha)} < \frac{m}{\gcd(m,b^{\alpha+1})}$.

For all the above mentioned reasons, the study of the state complexity of b-recognizable sets deserves special interest. In the present work, we propose ourselves to initiate a study of the state complexity of sets of the form $mX + r$, for any recognizable subset X of $\mathbb{N}$ (with respect to a given numeration system), any multiple m and any remainder r. In doing so, we aim at generalizing the previous framework concerning the case $X = \mathbb{N}$ only. Our study starts with the Thue-Morse set $\mathcal{T}$ of the so-called *evil numbers* [2], i.e. the natural numbers whose base-2 expansions contain an even number of occurrences of the digit 1. The characteristic sequence of this set corresponds to the ubiquitous Thue-Morse word $0110100110010110\cdots$, which is the fixed point starting with 0 of the morphism $0 \mapsto 01, 1 \mapsto 10$. This infinite word is one of the archetypical aperiodic automatic words, see the surveys [6, 34]. Many number-theoretic works devoted to sets of integers defined thanks to the Thue-Morse word exist, such as the study of additive and multiplicative properties, or iterations and sums of such sets [3, 12, 30]. In this vein, the set $\mathcal{T}$ seems to be a natural candidate to start with. The goal of this work is to provide a complete characterization of the minimal automata recognizing the sets $m\mathcal{T} + r$ for any multiple m and remainder r, and any base b which is a power of 2 (other bases are not relevant with the choice of the Thue-Morse set in view of Cobham's theorem). A previous work dealing with the case $r = 0$ recently appeared as a conference paper [14]. Surprisingly, the techniques of the present work, in particular the description of the left quotients (i.e. the states of the minimal automaton), are quite different from those we developed for the case $r = 0$.

This paper has the following organization. In Section 2, we recall the background that is necessary to tackle our problem. In Section 3, we state our main result and present the method that will be carried out for its proof. More precisely, we present the steps of our

construction of the minimal automaton accepting the base-2^p expansions of the elements of $m\mathcal{T} + r$ for any positive integers p and m, and any remainder $r \in \{0, \dots, m-1\}$. In Section 4, we give the details of the construction of the intermediate automata. In particular, we study the transitions of each automaton. Thus, at the end of Section 4, we are provided with an automaton recognizing the desired language. Then in Section 5, we study the properties of the built automata that will be needed for proving the announced state complexity result. The minimization procedure of the last automaton is handled in Section 6. This part is the most technical one and it deeply relies on the properties of the intermediate automata proved in the previous sections. In Section 7, we explicitly give the correspondence between the description of the minimal automaton recognizing $m\mathcal{T}$ obtained in [14] and that given in the present work in the particular case where $r = 0$. In Section 8, we show that the minimal automaton recognizing $m\overline{\mathcal{T}} + r$, where $\overline{\mathcal{T}}$ is the complement of the Thue-Morse set $\mathcal{T}$, is obtained directly from the one recognizing $m\mathcal{T} + r$ by moving the initial state. As a consequence, the state complexities of $m\overline{\mathcal{T}} + r$ and $m\mathcal{T} + r$ coincide. Finally, in Section 9, we discuss future work and give two related open problems.

2 Basics

In this text, we use the usual definitions and notation (alphabet, letter, word, language, free monoid, automaton, etc.) of formal language theory; for example, see [27, 36].

Nevertheless, let us give a few definitions and properties that will be central in this work. The empty word is denoted by ε. For a finite word w, $|w|$ designates its length and $|w|_a$ the number of occurrences of the letter a in w. A *regular language* is a language which is accepted by a finite automaton. For $L \subseteq A^*$ and $w \in A^*$, the *(left) quotient* of L by w is the language

$$w^{-1}L = \{u \in A^* : wu \in L\}.$$

As is well known, a language L over an alphabet A is regular if and only if it has finitely many quotients, that is, the set of languages

$$\{w^{-1}L : w \in A^*\}$$

is finite. The *state complexity* of a regular language is the number of its quotients: $\mathrm{Card}(\{w^{-1}L : w \in A^*\})$. It corresponds to the number of states of its minimal automaton. The following characterization of minimal automata will be used several times in this work: a deterministic finite automaton (or DFA for short) is minimal if and only if it is complete, reduced and accessible. A DFA is said to be *complete* if the transition function is total (i.e. transitions labeled with all possible letters start from every state), *reduced* if languages accepted from distinct states are distinct and *accessible* if every state can be reached from the initial state. The language accepted from a state q is denoted by L_q. Thus, the language accepted by a DFA is the language accepted from its initial state.

In what follows we will need a notion that is somewhat stronger than that of reduced DFAs. We say that a DFA has *disjoint states* if the languages accepted from distinct

states are disjoint: for distinct states p and q, we have $L_p \cap L_q = \varnothing$. A state q is said to be *coaccessible* if $L_q \neq \varnothing$ and, by extension, an automaton is *coaccessible* if all its states are coaccessible. Thus, any coaccessible DFA having disjoint states is reduced.

Now, let us give some background on numeration systems. Let $b \in \mathbb{N}_{\geqslant 2}$. We define A_b to be the alphabet $\{0, \ldots, b-1\}$. Elements of A_b are called *digits*. The number b is called the *base* of the numeration. In what follows we will make no distinction between a digit c in A_b and its *value* c in $[\![0, b-1]\!]$. Otherwise stated, we identify the alphabet A_b and the interval of integers $[\![0, b-1]\!]$. Note that here and throughout the text, we use the notation $[\![m, n]\!]$ to designate the interval of integers $\{m, m+1, \ldots, n\}$. The *b-expansion* of a positive integer n, which is denoted by $\mathrm{rep}_b(n)$, is the finite word $c_\ell \cdots c_0$ over A_b defined by

$$ n = \sum_{j=0}^{\ell} c_j b^j, \quad c_\ell \neq 0. $$

The *b-expansion* of 0 is the empty word: $\mathrm{rep}_b(0) = \varepsilon$. Conversely, for a word $w = c_\ell \cdots c_0$ over A_b, we write $\mathrm{val}_b(w) = \sum_{j=0}^{\ell} c_j b^j$. Thus we have $\mathrm{rep}_b \colon \mathbb{N} \to A_b^*$ and $\mathrm{val}_b \colon A_b^* \to \mathbb{N}$. Clearly, the function $\mathrm{val}_b \circ \mathrm{rep}_b$ is the identity from $\mathbb{N}$ to $\mathbb{N}$. Moreover, for any $w \in A_b^*$, the words $\mathrm{rep}_b(\mathrm{val}_b(w))$ and w only differ by the potential leading zeroes in w. Also note that for all subsets X of $\mathbb{N}$, we have $\mathrm{val}_b^{-1}(X) = 0^* \mathrm{rep}_b(X)$. A subset X of $\mathbb{N}$ is said to be *b-recognizable* if the language $\mathrm{rep}_b(X)$ is regular, or equivalently, if the language $\mathrm{val}_b^{-1}(X)$ is regular. In what follows, we will always consider automata accepting $\mathrm{val}_b^{-1}(X)$ instead of $\mathrm{rep}_b(X)$. The *state complexity* of a b-recognizable subset X of $\mathbb{N}$ *with respect to the base b* is the state complexity of the language $\mathrm{val}_b^{-1}(X)$. Note that the b-expansions are read from left to right, i.e. most significant digit first.

We will need to represent not only natural numbers, but also pairs of natural numbers. If $u = u_1 \cdots u_n \in A^*$ and $v = v_1 \cdots v_n \in B^*$ are words of the same length n, then we use the notation (u, v) to designate the word $(u_1, v_1) \cdots (u_n, v_n)$ of length n over the alphabet $A \times B$:

$$ (u, v) = (u_1, v_1) \cdots (u_n, v_n) \in (A \times B)^*. $$

For $(m, n) \in \mathbb{N}^2$, we write

$$ \mathrm{rep}_b(m, n) = (0^{\ell - |\mathrm{rep}_b(m)|} \mathrm{rep}_b(m), 0^{\ell - |\mathrm{rep}_b(n)|} \mathrm{rep}_b(n)) $$

where $\ell = \max\{|\mathrm{rep}_b(m)|, |\mathrm{rep}_b(n)|\}$. Otherwise stated, we add leading zeroes to the shortest expansion (if any) in order to obtain two words of the same length. Finally, for a subset X of $\mathbb{N}^2$, we write

$$ \mathrm{val}_b^{-1}(X) = (0, 0)^* \mathrm{rep}_b(X). $$

3 Method

The Thue-Morse set, which we denote by $\mathcal{T}$, is the set of all natural numbers whose base-2 expansions contain an even number of occurrences of the digit 1:

$$ \mathcal{T} = \{n \in \mathbb{N} \colon |\mathrm{rep}_2(n)|_1 \in 2\mathbb{N}\}. $$

The Thue-Morse set $\mathcal{T}$ is 2-recognizable since the language $\mathrm{val}_2^{-1}(\mathcal{T})$ is accepted by the automaton depicted in Figure 1. More precisely, the Thue-Morse set $\mathcal{T}$ is 2^p-recognizable
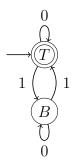

Figure 1: The Thue-Morse set is 2-recognizable.

for all $p \in \mathbb{N}_{\geqslant 1}$ and is not b-recognizable for any other base b. This is a consequence of the famous theorem of Cobham. Two positive integers are said to be *multiplicatively independent* if their only common integer power is 1 and are said to be *multiplicatively dependent* otherwise.

Theorem 1 (Cobham [19]).

- *Let b, b' be two multiplicatively independent bases. Then a subset of $\mathbb{N}$ is both b-recognizable and b'-recognizable if and only if it is a finite union of arithmetic progressions.*

- *Let b, b' be two multiplicatively dependent bases. Then a subset of $\mathbb{N}$ is b-recognizable if and only if it is b'-recognizable.*

In the case of the Thue-Morse set, it is easily seen that, for each $p \in \mathbb{N}_{\geqslant 1}$, the language $\mathrm{val}_{2^p}^{-1}(\mathcal{T})$ is accepted by the DFA $(\{T, B\}, T, T, A_{2^p}, \delta)$ where for all $X \in \{T, B\}$ and all $a \in A_{2^p}$,

$$\delta(X, a) = \begin{cases} X & \text{if } a \in \mathcal{T} \\ \overline{X} & \text{otherwise} \end{cases}$$

where $\overline{T} = B$ and $\overline{B} = T$. For example this automaton is depicted in Figure 2 for $p = 2$.

In order to avoid a systematic case separation, we introduce the following notation: for $X \in \{T, B\}$ and $n \in \mathbb{N}$, we define

$$X_n = \begin{cases} X & \text{if } n \in \mathcal{T} \\ \overline{X} & \text{otherwise.} \end{cases}$$

With this notation, we can simply rewrite the definition of the transition function δ as $\delta(X, a) = X_a$.

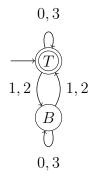The following proposition is well known; for example see [11].

Figure 2: The Thue-Morse set is 4-recognizable.

Proposition 2. *Let $b \in \mathbb{N}_{\geqslant 2}$ and $m, t \in \mathbb{N}$. If X is b-recognizable, then so is $mX + t$. Otherwise stated, the dilation $n \mapsto mn$ and translation $n \mapsto n+t$ preserve b-recognizability.*

In particular, for any $m, t \in \mathbb{N}$ and $p \in \mathbb{N}_{\geqslant 1}$, the set $m\mathcal{T} + t$ is 2^p-recognizable. The aim of this work is to show the following result.

Theorem 3. *Let m, p be positive integers and $r \in [\![0, m-1]\!]$. Then the state complexity of $m\mathcal{T} + r$ with respect to the base 2^p is equal to*

$$2k + \left\lceil \frac{z}{p} \right\rceil$$

if $m = k2^z$ with k odd.

Remark that the state complexity of $m\mathcal{T} + r$ (with respect to the base 2^p) is independent of the value of the remainder r.

Our proof of Theorem 3 is constructive. In order to describe the minimal DFA of $\mathrm{val}_{2^p}^{-1}(m\mathcal{T} + r)$, we will successively construct several automata. First, we build a DFA $\mathcal{A}_{\mathcal{T}, 2^p}$ accepting the language

$$\mathrm{val}_{2^p}^{-1}(\mathcal{T} \times \mathbb{N}).$$

Then we build a DFA $\mathcal{A}_{m,r,b}$ accepting the language

$$\mathrm{val}_b^{-1}\big(\{(n, mn + r) \colon n \in \mathbb{N}\}\big).$$

Note that we do the latter step for any integer base b and not only for powers of 2. Next, we consider the product automaton $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T}, 2^p}$. This DFA accepts the language

$$\mathrm{val}_{2^p}^{-1}\big(\{(t, mt + r) \colon t \in \mathcal{T}\}\big).$$

Finally, a finite automaton $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T}, 2^p})$ accepting $\mathrm{val}_{2^p}^{-1}(m\mathcal{T} + r)$ is obtained by projecting the label of each transition in $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T}, 2^p}$ onto its second component. At each step of our construction, we check that the automaton under consideration is minimal (and hence deterministic) and the ultimate step precisely consists in a minimization procedure.

From now on, we fix some positive integers m, p and some remainder $r \in [\![0, m-1]\!]$. We also let z and k be the unique integers such that $m = k2^z$ with k odd. Finally we let $R = |\mathrm{rep}_{2^p}(r)|$.

4 Construction of the intermediate automata

4.1 The automaton $\mathcal{A}_{\mathcal{T}, 2^p}$

First, we build a DFA $\mathcal{A}_{\mathcal{T}, 2^p}$ accepting the language $\mathrm{val}_{2^p}^{-1}(\mathcal{T} \times \mathbb{N})$. This DFA is a modified version of the automaton accepting $\mathrm{val}_{2^p}^{-1}(\mathcal{T})$ defined in the previous section. Namely, we replace each transition labeled by $a \in A_{2^p}$ by 2^p copies of itself labeled by (a, b), for each $b \in A_{2^p}$. Formally,

$$\mathcal{A}_{\mathcal{T}, 2^p} = (\{T, B\}, T, T, A_{2^p} \times A_{2^p}, \delta_{\mathcal{T}, 2^p})$$

where, for all $X \in \{T, B\}$ and all $a, b \in A_{2^p}$, we have $\delta_{\mathcal{T}, 2^p}(X, (a, b)) = X_a$. For example, the automata $\mathcal{A}_{\mathcal{T}, 2}$ and $\mathcal{A}_{\mathcal{T}, 4}$ are depicted in Figure 3.
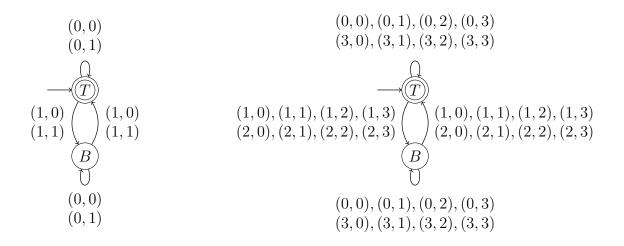

Figure 3: The automata $\mathcal{A}_{\mathcal{T}, 2}$ (left) and $\mathcal{A}_{\mathcal{T}, 4}$ (right).

Lemma 4. *Let* $u, v \in A_{2^p}^*$. *Then* $\mathrm{val}_{2^p}(uv) \in \mathcal{T}$ *if and only if, either* $\mathrm{val}_{2^p}(u) \in \mathcal{T}$ *and* $\mathrm{val}_{2^p}(v) \in \mathcal{T}$, *or* $\mathrm{val}_{2^p}(u) \notin \mathcal{T}$ *and* $\mathrm{val}_{2^p}(v) \notin \mathcal{T}$.

Proof. Let $\tau \colon A_{2^p}^* \to A_{2^p}^*$ be the p-uniform morphism defined by $\tau(a) = 0^{p - |\mathrm{rep}_2(a)|}\mathrm{rep}_2(a)$ for each $a \in A_{2^p}$. Then, for all $w \in A_{2^p}^*$, we have $\mathrm{val}_{2^p}(w) = \mathrm{val}_2(\tau(w))$. Therefore, $\mathrm{val}_{2^p}(w) \in \mathcal{T}$ if and only if $|\tau(w)|_1 \in 2\mathbb{N}$. Since τ is a morphism, we have $|\tau(uv)|_1 = |\tau(u)|_1 + |\tau(v)|_1$. Hence $|\tau(uv)|_1$ is even if and only if $|\tau(u)|_1$ and $|\tau(v)|_1$ are both even or both odd. $\square$

Lemma 5. *For all $X \in \{T, B\}$ and $(u, v) \in (A_{2^p} \times A_{2^p})^*$, we have*

$$\delta_{\mathcal{T}, 2^p}(X, (u, v)) = X_{\mathrm{val}_{2^p}(u)}.$$

Proof. We do the proof by induction on $|(u, v)|$. The case $|(u, v)| = 0$ is trivial. Now let $X \in \{T, B\}$ and let $(ua, vb) \in (A_{2^p} \times A_{2^p})^*$ with $a, b \in A_{2^p}$. We suppose that the result is satisfied for (u, v) and we show that it is also true for (ua, vb). Let $Y = \delta_{\mathcal{T}, 2^p}(X, (u, v))$. By the induction hypothesis, we have $Y = X_{\mathrm{val}_{2^p}(u)}$. Thus we obtain

$$\delta_{\mathcal{T}, 2^p}(X, (ua, vb)) = \delta_{\mathcal{T}, 2^p}(Y, (a, b)) = Y_a = (X_{\mathrm{val}_{2^p}(u)})_a = X_{\mathrm{val}_{2^p}(ua)}$$

where we have used Lemma 4 for the last step. $\qquad\square$

4.2 The automaton $\mathcal{A}_{m,r,b}$

In this section, we consider an arbitrary integer base b. Let

$$\mathcal{A}_{m,r,b} = (\llbracket 0, m{-}1 \rrbracket, 0, r, A_b \times A_b, \delta_{m,r,b})$$

where the (partial) transition function $\delta_{m,r,b}$ is defined as follows: for $i, j \in \llbracket 0, m{-}1 \rrbracket$ and $d, e \in A_b$, we set

$$\delta_{m,r,b}(i, (d, e)) = j \qquad \Longleftrightarrow \qquad bi + e = md + j.$$

The DFA $\mathcal{A}_{m,r,b}$ accepts the language $\mathrm{val}_b^{-1}\big(\{(n, mn{+}r) \colon n \in \mathbb{N}\}\big)$. We refer the interested reader to [38]. For example, the automaton $\mathcal{A}_{6,2,4}$ is depicted in Figure 4.
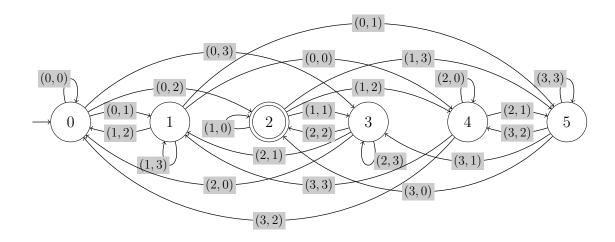

Figure 4: The automaton $\mathcal{A}_{6,2,4}$ accepts the language $\mathrm{val}_4^{-1}\big(\{(n, 6n + 2) \colon n \in \mathbb{N}\}\big)$.

Note that the automaton $\mathcal{A}_{m,r,b}$ is not complete (see Remark 6). Also note that there is always a loop labeled by $(0, 0)$ on the initial state 0.

Remark 6. For each $i \in [\![0, m{-}1]\!]$ and $e \in A_b$, there exist unique $d \in A_b$ and $j \in [\![0, m{-}1]\!]$ such that $\delta_{m,r,b}(i, (d, e)) = j$. Indeed, d and j are unique since they are the quotient and remainder of the Euclidean division of $bi + e$ by m. We still have to check that $d < b$. We have
$$bi + e = md + j \iff d = \frac{bi + e - j}{m}.$$
Since $i \leqslant m{-}1$, $j \geqslant 0$ and $e < b$, we have
$$\frac{bi + e - j}{m} < \frac{b(m{-}1) + b}{m} = b.$$

Lemma 7. *For* $i, j \in [\![0, m{-}1]\!]$ *and* $(u, v) \in (A_b \times A_b)^*$, *we have*
$$\delta_{m,r,b}(i, (u, v)) = j \iff b^{|(u,v)|} i + \mathrm{val}_b(v) = m \, \mathrm{val}_b(u) + j.$$

Proof. We do the proof by induction on $n = |(u, v)|$. If n is equal to 0, the result is clear. Now let $i, j \in [\![0, m{-}1]\!]$ and let $(du, ev) \in (A_b \times A_b)^*$ with $d, e \in A_b$ and $|(u, v)| = n$. We suppose that the result is satisfied for (u, v) and we show that it is also true for (du, ev). We use the notation $\mathrm{DIV}(x, y)$ and $\mathrm{MOD}(x, y)$ to designate the quotient and the remainder of the Euclidean division of x by y (thus, we have $\mathrm{DIV}(x, y) = \lfloor \frac{x}{y} \rfloor$). By definition of the transition function, we have
$$\delta_{m,r,b}(i, (du, ev)) = j \iff d = \mathrm{DIV}(bi + e, m) \ \text{ and } \ \delta_{m,r,b}(\mathrm{MOD}(bi + e, m), (u, v)) = j.$$
By using the induction hypothesis, we have
$$\begin{aligned} \delta_{m,r,b}(bi + e - md, (u, v)) = j &\iff b^n (bi + e - md) + \mathrm{val}_b(v) = m \, \mathrm{val}_b(u) + j \\ &\iff b^{n+1} i + \mathrm{val}_b(ev) = m \, \mathrm{val}_b(du) + j. \end{aligned}$$
To be able to conclude the proof, we still have to show that
$$b^{n+1} i + \mathrm{val}_b(ev) = m \, \mathrm{val}_b(du) + j \tag{1}$$
implies
$$d = \mathrm{DIV}(bi + e, m).$$
Thus, suppose that (1) is true. Then
$$b^{n+1} i + b^n e + \mathrm{val}_b(v) = m(b^n d + \mathrm{val}_b(u)) + j.$$
Since $\mathrm{val}_b(u)$ and $\mathrm{val}_b(v)$ are less than b^n, $d \geqslant 0$, $j < m$ and $b^n d + \mathrm{val}_b(u) \geqslant 0$, we obtain
$$\begin{aligned} d &= \mathrm{DIV}(b^n d + \mathrm{val}_b(u), b^n) \\ &= \mathrm{DIV}(\mathrm{DIV}(b^{n+1} i + b^n e + \mathrm{val}_b(v), m), b^n) \\ &= \mathrm{DIV}(\mathrm{DIV}(b^{n+1} i + b^n e + \mathrm{val}_b(v), b^n), m) \\ &= \mathrm{DIV}(b \, i + e, m) \end{aligned}$$
as desired. $\qquad\square$

Remark 8. It is easily checked that Remark 6 extends from letters to words: for each $i \in [\![0, m{-}1]\!]$ and $v \in A_b^*$, there exist unique $u \in A_b^*$ and $j \in [\![0, m{-}1]\!]$ such that $\delta_{m,r,b}(i, (u, v)) = j$. In particular, the word u must have the same length as the word v, and hence $\mathrm{val}_b(u) < b^{|v|}$.

4.3 The projected automaton $\Pi(\mathcal{A}_{m,r,b})$

In this section again, b is an arbitrary integer base. We consider the automaton obtained by projecting the label of each transition of $\mathcal{A}_{m,r,b}$ onto its second component. We denote by $\Pi(\mathcal{A}_{m,r,b})$ the automaton obtained thanks to this projection. Thanks to Remark 6, the automaton $\Pi(\mathcal{A}_{m,r,b})$ is deterministic and complete. We denote by $\delta^{\Pi}_{m,r,b}$ the corresponding transition function. For example, the automaton $\Pi(\mathcal{A}_{6,2,4})$ is depicted in Figure 5.
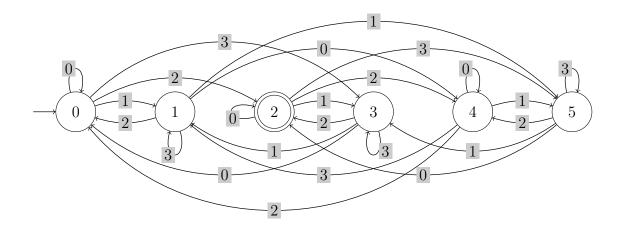

Figure 5: The projected automaton $\Pi(\mathcal{A}_{6,2,4})$.

Note that the automaton $\Pi(\mathcal{A}_{m,r,b})$ actually corresponds to the "classical" construction of an automaton recognizing $m\mathbb{N} + r$ in base b, see for example [1].

Lemma 9. *For $i, j \in [\![0, m-1]\!]$ and $v \in A_b^*$, we have*

$$\delta^{\Pi}_{m,r,b}(i, v) = j \iff b^{|v|} i + \mathrm{val}_b(v) \equiv j \pmod{m}.$$

Proof. Let $i, j \in [\![0, m-1]\!]$ and $v \in A_b^*$. If $\delta^{\Pi}_{m,r,b}(i, v) = j$, then there exists a word u of the same length as v such that $\delta_{m,r,b}(i, (u, v)) = j$. By Lemma 7, we get $b^{|v|} i + \mathrm{val}_b(v) \equiv j$ (mod m). Conversely, suppose that there exists some $\ell \in \mathbb{N}$ such that $b^{|v|} i + \mathrm{val}_b(v) = m\ell + j$. Since $i \leqslant m-1$, $\mathrm{val}_b(v) < b^{|v|}$ and $j \geqslant 0$, we necessarily have

$$\ell = \frac{b^{|v|}i + \mathrm{val}_b(v) - j}{m} < \frac{b^{|v|}(m-1) + b^{|v|}}{m} = b^{|v|}.$$

Hence $|\mathrm{rep}_b(\ell)| \leqslant |v|$ and the word $u = 0^{|v|-|\mathrm{rep}_b(\ell)|}\mathrm{rep}_b(\ell)$ has length $|v|$ and is such that $\mathrm{val}_b(u) = \ell$. The conclusion follows from Lemma 7. $\qquad\square$

4.4 The product automaton $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$

In this section, we study the product automaton $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$. Since the states of $\mathcal{A}_{m,r,2^p}$ are numbered from 0 to $m-1$ and those of $\mathcal{A}_{\mathcal{T},2^p}$ are T and B, we denote the states of

the product automaton by

$$(0, T), \ldots, (m{-}1, T) \text{ and } (0, B), \ldots, (m{-}1, B).$$

The transitions of $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$ are defined as follows. For $i, j \in [\![0, m{-}1]\!]$, $X, Y \in \{T, B\}$ and $d, e \in A_{2^p}$, there is a transition labeled by (d, e) from the state (i, X) to the state (j, Y) if and only if

$$2^p i + e = md + j \quad \text{and} \quad Y = X_d.$$

We denote by $\delta_\times$ the (partial) transition function of this product automaton. The initial state is $(0, T)$ and the only final state is (r, T).

Lemma 10. *For all $i, j \in [\![0, m{-}1]\!]$, $X, Y \in \{T, B\}$ and $(u, v) \in (A_{2^p} \times A_{2^p})^*$, we have $\delta_\times((i, X), (u, v)) = (j, Y)$ if and only if*

$$2^{p\,|(u,v)|}\, i + \mathrm{val}_{2^p}(v) = m\, \mathrm{val}_{2^p}(u) + j \quad \text{and} \quad Y = X_{\mathrm{val}_{2^p}(u)}.$$

Proof. It suffices to combine Lemmas 5 and 7. $\qquad\square$

In Figure 6, we have depicted the automaton $\mathcal{A}_{6,2,4} \times \mathcal{A}_{\mathcal{T},4}$, as well as the automata $\mathcal{A}_{6,2,4}$ and $\mathcal{A}_{\mathcal{T},4}$, which we have placed in such a way that the labels of the product automaton can be easily deduced. Here and in the next figures, states are named iX instead of (i, X) for clarity.

4.5 The projection $\Pi\left(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}\right)$ of the product automaton

Now, we provide a DFA accepting the language $\mathrm{val}_{2^p}^{-1}(m\mathcal{T} + r)$. This automaton is denoted by $\Pi\left(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}\right)$ and is defined from the automaton $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$ by only keeping the second component of the label of each transition. Formally, the states of $\Pi\left(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}\right)$ are

$$(0, T), \ldots, (m{-}1, T) \text{ and } (0, B), \ldots, (m{-}1, B),$$

the initial state is $(0, T)$, the only final state is (r, T), and the transitions are defined as follows. For $i, j \in [\![0, m{-}1]\!]$, $X, Y \in \{T, B\}$ and $e \in A_{2^p}$, there is a transition labeled by e from the state (i, X) to the state (j, Y) if and only if there exists $d \in A_{2^p}$ such that

$$2^p i + e = md + j \quad \text{and} \quad Y = X_d.$$

We denote by $\delta_\times^\Pi$ the (partial) transition function of this product automaton.
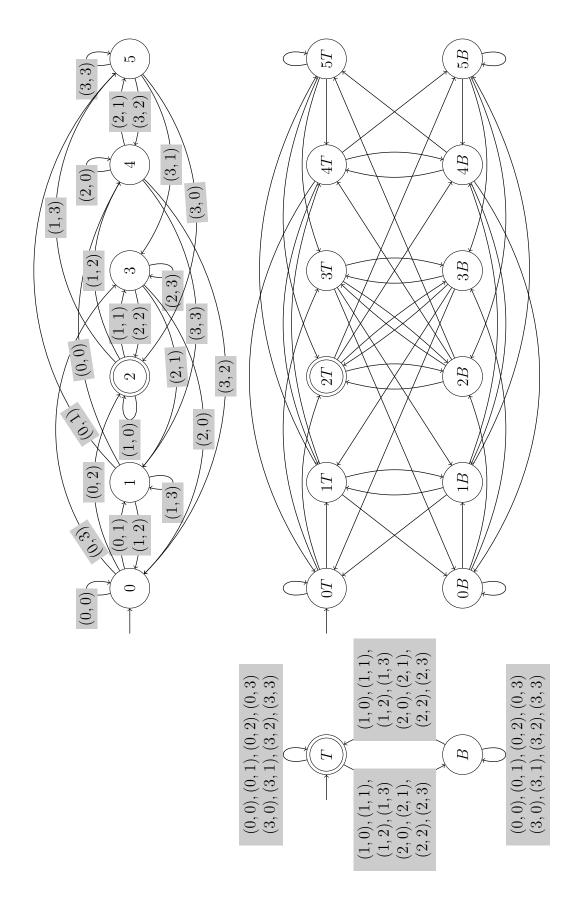
Example 11. The automata $\mathcal{A}_{6,2,4} \times \mathcal{A}_{\mathcal{T},4}$ and $\Pi\left(\mathcal{A}_{6,2,4} \times \mathcal{A}_{\mathcal{T},4}\right)$ are depicted in Figures 6 and 7 respectively. In Figure 7, all edges labeled by 0 (1, 2 and 3 respectively) are represented in black (blue, red and green respectively).

Lemma 12. *For all $i, j \in [\![0, m{-}1]\!]$, $X, Y \in \{T, B\}$ and $v \in A_{2^p}^*$, we have $\delta_\times^\Pi((i, X), v) = (j, Y)$ if and only if there exists $\ell \in \mathbb{N}$ such that*

$$2^{p\,|v|}\, i + \mathrm{val}_{2^p}(v) = m\ell + j \quad \text{and} \quad Y = X_\ell.$$

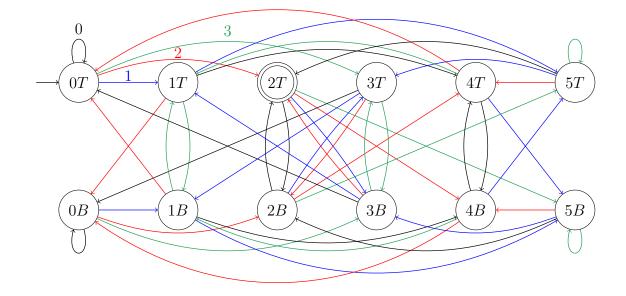Figure 6: The product automaton $\mathcal{A}_{6,2,4} \times \mathcal{A}_{\mathcal{T},4}$

Figure 7: The projected automaton $\Pi\left(\mathcal{A}_{6,2,4} \times \mathcal{A}_{\mathcal{T},4}\right)$.

Proof. We have $\delta_{\times}^{\Pi}((i, X), v) = (j, Y)$ if and only if there exists some word u over A_{2^p} of the same length as v such that $\delta_{\times}((i, X), (u, v)) = (j, Y)$. Take $\ell = \mathrm{val}_{2^p}(u)$. The conclusion follows from Lemma 10 and a similar argument as in the proof of Lemma 9. $\qquad\square$

Remark 13. Note that in the statement of Lemma 12, the integer ℓ is necessarily less than $2^{p|v|}$. This is due to the fact that if v is the label of some path in the projected automaton $\Pi\left(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}\right)$, there must be a word u of the same length ℓ as v such that the pair (u, v) is the label of a path in the automaton $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$. This can be deduced directly from the computation: if $2^{p|v|} i + \mathrm{val}_{2^p}(v) = m\ell + j$ (with $i, j \in [\![0, m-1]\!]$) then

$$\ell = \frac{2^{p|v|} i + \mathrm{val}_{2^p}(v) - j}{m} < \frac{2^{p|v|}(i+1)}{m} \leqslant 2^{p|v|}.$$

5 Properties of the intermediate automata

Now we prove some properties of the automata $\mathcal{A}_{\mathcal{T},2^p}$, $\mathcal{A}_{m,r,b}$, $\Pi(\mathcal{A}_{m,r,b})$, $\mathcal{A}_{\mathcal{T},2^p} \times \mathcal{A}_{m,r,b}$ and $\Pi(\mathcal{A}_{\mathcal{T},2^p} \times \mathcal{A}_{m,r,b})$ that will be useful for our concerns.

5.1 Properties of $\mathcal{A}_{\mathcal{T},2^p}$

Lemma 14. *For all $X, Y \in \{T, B\}$ and $(u, v) \in (A_{2^p} \times A_{2^p})^*$, we have*

$$\delta_{\mathcal{T},2^p}(X, (u, v)) = Y \iff \delta_{\mathcal{T},2^p}(\overline{X}, (u, v)) = \overline{Y}.$$

Proof. This directly follows from Lemma 5. $\qquad\square$

Proposition 15. *The automaton $\mathcal{A}_{\mathcal{T},2^p}$ is complete, accessible, coaccessible and has disjoint states. In particular, it is the minimal automaton of $\mathrm{val}_{2^p}^{-1}(\mathcal{T} \times \mathbb{N})$.*

Proof. These properties are all straightforward verifications. $\square$

5.2 Properties of $\Pi(\mathcal{A}_{m,r,2^p})$ and $\mathcal{A}_{m,r,2^p}$

In what follows, we let $K = |\mathrm{rep}_{2^p}((k-1)2^z)|$, provided that $k > 1$. Then we define a permutation σ of the integers in $[\![0, k-1]\!]$ by $\sigma(i) = -2^{pK-z}i \bmod k$. Note that σ permutes the integers $0, 1, \ldots, k-1$ because k is odd. Further, we define w_i to be the unique word of length K representing $\sigma(i)2^z$ in base 2^p:

$$w_i = 0^{K-|\mathrm{rep}_{2^p}(\sigma(i)2^z)|}\mathrm{rep}_{2^p}(\sigma(i)2^z)$$

for each $i \in [\![0, k-1]\!]$. Note that the words w_i are well defined since, by the choice of K, we have $\sigma(i)2^z \leqslant (k-1)2^z < 2^{pK}$ for every $i \in [\![0, k-1]\!]$.

Lemma 16. *If $k > 1$ then $pK \geqslant z$.*

Proof. We have

$$K = \left\lfloor \log_{2^p}\left((k-1)2^z\right)\right\rfloor + 1 = \left\lfloor \log_{2^p}(k-1) + \frac{z}{p}\right\rfloor + 1 \geqslant \left\lfloor \frac{z}{p}\right\rfloor + 1 \geqslant \left\lceil \frac{z}{p}\right\rceil.$$

Thus $pK \geqslant p\left\lceil \frac{z}{p}\right\rceil \geqslant z$. $\square$

Lemma 17. *Suppose that $k > 1$ and let $i \in [\![0, k-1]\!]$. Then the word w_i leads from the state i to the state 0 in the automaton $\Pi(\mathcal{A}_{m,r,2^p})$. Otherwise stated, $\delta_{m,r,2^p}^{\Pi}(i, w_i) = 0$.*

Proof. The word w_i has length K and from Lemma 16, we know that $pK \geqslant z$. By Lemma 9, we have

$$
\begin{aligned}
\delta_{m,r,2^p}^{\Pi}(i, w_i) = 0 &\iff 2^{pK}i + \mathrm{val}_{2^p}(w_i) \equiv 0 \pmod{m} \\
&\iff 2^{pK}i + \sigma(i)2^z \equiv 0 \pmod{k2^z} \\
&\iff 2^{pK-z}i + \sigma(i) \equiv 0 \pmod{k}.
\end{aligned}
$$

The result follows from the definition of σ. $\square$

Lemma 18. *Suppose that $k > 1$ and let $i, j, \in [\![0, k-1]\!]$. For any $\ell \in \mathbb{N}$, the word*

$$w_i(\mathrm{rep}_{2^p}(m))^{\ell}\mathrm{rep}_{2^p}(r)$$

is accepted from j in the automaton $\Pi(\mathcal{A}_{m,r,2^p})$ if and only if $i = j$.

Proof. Let $\ell \in \mathbb{N}$ and, for each $i \in [\![0, k-1]\!]$, let $y_i = w_i(\text{rep}_{2^p}(m))^\ell \text{rep}_{2^p}(r)$. Recall that we have set $R = |\text{rep}_{2^p}(r)|$. Further, set $M = |\text{rep}_{2^p}(m)|$. Then $|y_i| = K + \ell M + R$ and from Lemma 16, we know that $pK \geqslant z$. Therefore, we have

$$j2^{p|y_i|} + \text{val}_{2^p}(y_i) \equiv r \pmod{m}$$

$$\Longleftrightarrow j2^{p(K+\ell M+R)} + \text{val}_{2^p}(w_i)2^{p(\ell M+R)} + \sum_{s=0}^{\ell-1} m2^{p(sM+R)} + r \equiv r \pmod{m}$$

$$\Longleftrightarrow j2^{p(K+\ell M+R)} + \sigma(i)2^{z+p(\ell M+R)} \equiv 0 \pmod{k2^z}$$

$$\Longleftrightarrow j2^{p(K+\ell M+R)-z} + \sigma(i)2^{p(\ell M+R)} \equiv 0 \pmod{k}$$

$$\Longleftrightarrow j2^{p(K+\ell M+R)-z} - i2^{p(K+\ell M+R)-z} \equiv 0 \pmod{k}$$

$$\Longleftrightarrow j \equiv i \pmod{k}$$

$$\Longleftrightarrow j = i.$$

The conclusion follows from Lemma 9. $\square$

Proposition 19. *The automaton* $\Pi(\mathcal{A}_{m,r,2^p})$ *is complete, accessible and coaccessible.*

Proof. By Lemma 9, we have $\delta_{m,r,2^p}^\Pi(0, \text{rep}_{2^p}(i)) = i$ for all $i \in [\![0, m-1]\!]$. Thus, the automaton $\Pi(\mathcal{A}_{m,r,2^p})$ is accessible and in particular, we have $\delta_{m,r,2^p}^\Pi(0, \text{rep}_{2^p}(r)) = r$. By Lemma 17, we have $\delta_{m,r,2^p}^\Pi(i, w_i) = 0$ for all $i \in [\![0, m-1]\!]$, and hence, the concatenation $w_i\text{rep}_{2^p}(r)$ leads from i to r. This shows that $\Pi(\mathcal{A}_{m,r,2^p})$ is also coaccessible. In order to see that it is complete, observe that for every state $i \in [\![0, m-1]\!]$ and every digit $e \in A_{2^p}$, there is a transition labeled by e from i to the state $2^p i + e \bmod m$ (see Remark 6). $\square$

In [14], we proved the coaccessibility for arbitrary integer bases b by using a different method. In the present case of a base which is a power of two, the argument is more straightforward since we are able to explicitly provide a word that is accepted from any state i.

The automaton $\Pi(\mathcal{A}_{m,r,b})$ is not minimal in general: it is minimal if and only if m and b are coprime; see for example [1].

Proposition 20. *The automaton* $\mathcal{A}_{m,r,2^p}$ *is accessible, coaccessible and has disjoint states.*

Proof. It directly follows from Proposition 19 that $\mathcal{A}_{m,r,2^p}$ is accessible and coaccessible. Now, let $i, j \in [\![0, m-1]\!]$ and let $(u, v) \in L_i \cap L_j$. By Lemma 7, we have

$$2^{p|(u,v)|}i + \text{val}_b(v) = m\,\text{val}_{2^p}(u) + r \quad \text{and} \quad 2^{p|(u,v)|}j + \text{val}_b(v) = m\,\text{val}_{2^p}(u) + r,$$

which implies that $i = j$. This proves that $\mathcal{A}_{m,r,2^p}$ has disjoint states. $\square$

In a reduced DFA, there can be at most one non-coaccessible state. Thus, we deduce from Proposition 20 that $\mathcal{A}_{m,r,2^p}$ is indeed the *trim minimal* automaton of the language $\text{val}_{2^p}^{-1}(\{(n, mn + r) : n \in \mathbb{N}\})$, that is the automaton obtained by removing the only non-coaccessible state from its minimal automaton.

5.3 Properties of $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$

Lemma 21. *For all $i, j \in [\![0, m-1]\!]$, $X, Y \in \{T, B\}$ and $(u, v) \in (A_{2^p} \times A_{2^p})^*$, we have*

$$\delta_{\mathcal{T},2^p}((i, X), (u, v)) = (j, Y) \iff \delta_{\mathcal{T},2^p}((i, \overline{X}), (u, v)) = (j, \overline{Y}).$$

Proof. This directly follows from Lemma 14. $\qquad\square$

Lemma 22. *Let $i \in [\![0, m-1]\!]$ and $X \in \{T, B\}$. The word $\mathrm{rep}_{2^p}(1, m)$ is the label of a path from the state $(0, X)$ to the state $(0, \overline{X})$ in $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$.*

Proof. This directly follows from Lemma 10. $\qquad\square$

Proposition 23. *The automaton $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$ accepts $\mathrm{val}_{2^p}^{-1}(\{(t, mt + r) \colon t \in \mathcal{T}\})$, is accessible, coaccessible and has disjoint states.*

Proof. By construction of the product automaton and since

$$\{(n, mn + r) \colon n \in \mathbb{N}\} \cap (\mathcal{T} \times \mathbb{N}) = \{(t, mt + r) \colon t \in \mathcal{T}\},$$

we get that the product automaton $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$ accepts the language

$$\mathrm{val}_{2^p}^{-1}(\{(t, mt + r) \colon t \in \mathcal{T}\}).$$

By Lemma 10, we can check that for every $i \in [\![0, m-1]\!]$, the states (i, T) and (i, B) are accessible thanks to the word $\mathrm{rep}_{2^p}(0, i)$ and $\mathrm{rep}_{2^p}(1, m + i)$ respectively. Hence, $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$ is accessible.

To show the coaccessibility, we now fix some $i \in [\![0, m-1]\!]$ and $X \in \{T, B\}$. By Lemma 17, we know that there exists a word w_i that leads from the state i to the state 0 in the automaton $\Pi(\mathcal{A}_{m,r,2^p})$. Thus, there exists a word u of the same length as w_i such that the word (u, w_i) leads from i to 0 in $\mathcal{A}_{m,r,2^p}$. Now, by reading (u, w_i) from (i, X) in $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$, we reach either the state $(0, T)$ or the state $(0, B)$. If we reach $(0, T)$, then the concatenation $(u, w_i)\mathrm{rep}_b(0, r)$ leads from the state (i, X) to (r, T) in $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$. If we reach $(0, B)$ instead, then we may apply Lemma 22 in order to obtain that the concatenation $(u, w_i)\mathrm{rep}_b(1, m)\mathrm{rep}_b(0, r)$ leads from (i, X) to (r, T) in $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$. This proves that $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$ is coaccessible.

The fact that $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$ has disjoint states follows from Propositions 15 and 20. $\qquad\square$

5.4 Properties of $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$

Lemma 24. *For all $i, j \in [\![0, m-1]\!]$, $X, Y \in \{T, B\}$ and $v \in A_{2^p}^*$, we have*

$$\delta_\times^\Pi((i, X), v) = (j, Y) \implies \delta_{m,r,2^p}^\Pi(i, v) = j.$$

Proof. This is a direct verification. $\qquad\square$

Lemma 25. *For every $i \in [\![0, m-1]\!]$, the states (i, T) and (i, B) are disjoint in the projected automaton $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$.*

Proof. Proceed by contradiction and suppose that a word v over A_{2^p} is accepted from both (i, T) and (i, B) in $\Pi\left(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}\right)$ for some $i \in [\![0, m-1]\!]$. Then there exists words u and u' over A_{2^p} of length $|v|$ such that, in $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$, the words (u, v) and (u', v) are accepted from (i, T) and (i, B) respectively. But from Remark 8, we must have $u = u'$. Hence the word (u, v) is accepted from both (i, T) and (i, B) in $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$, contradicting that this automaton has disjoint states (see Proposition 23). $\qquad\square$

Proposition 26. *The automaton $\Pi\left(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}\right)$ accepts $\mathrm{val}_{2^p}^{-1}(m\mathcal{T}+r)$, is deterministic, complete, accessible and coaccessible.*

Proof. By construction, $\Pi\left(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}\right)$ accepts $\mathrm{val}_{2^p}^{-1}(m\mathcal{T}+r)$; see Section 3. The fact that this automaton is deterministic and complete follows from Remark 6. It is accessible and coaccessible because $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$ is. $\qquad\square$

6 Minimization of $\Pi\left(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}\right)$

We start by defining some classes of states of $\Pi\left(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}\right)$. Our aim is twofold. First, we will prove that those classes consist in *indistinguishable* states, i.e. accepting the same language. Second, we will show that states belonging to different classes are *distinguishable*, i.e. accept different languages. Otherwise stated, these classes correspond to the left quotients $w^{-1}L$ where w is any finite word over the alphabet A_{2^p} and $L = \mathrm{val}_{2^p}^{-1}(m\mathcal{T} + r)$.

6.1 Definition of the classes

Recall that $R = |\mathrm{rep}_{2^p}(r)|$. We define $N = \max\{\lceil \frac{z}{p} \rceil, R\}$. The classes we are going to define are closely related to the base 2^p-expansion of the remainder r with some additional leading zeroes. More precisely, we have to consider the word $0^{N-R}\mathrm{rep}_{2^p}(r)$, which is the unique word over the alphabet A_{2^p} with length N and 2^p-value r. This word is equal to the 2^p-expansion $\mathrm{rep}_{2^p}(r)$ if and only if $N = R$, i.e. $\lceil \frac{z}{p} \rceil \leqslant R$.

Definition 27. For $\alpha \in [\![0, N]\!]$, we define

$$
C'_\alpha = \begin{cases} \left\{ \left(\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell \frac{m}{2^{p\alpha}}, T_\ell \right) : 0 \leqslant \ell \leqslant 2^{p\alpha}-1 \right\} & \text{if } \alpha \leqslant \frac{z}{p} \\ \left\{ \left(\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell k, T_\ell \right) : 0 \leqslant \ell \leqslant 2^z-1 \right\} & \text{if } \alpha \geqslant \frac{z}{p}. \end{cases}
$$

Note that in the case where z is divisible by p, the two cases of the definition coincide for the value $\alpha = \frac{z}{p}$.

Let us comment the previous definition, which may seem quite technical at first. The first elements of the sets C'_α are the integer part of the remainder r divided by increasing powers of the base 2^p, i.e. r divided by $2^{p\alpha}$ for the set indexed by α. The further elements of a set C'_α are obtained by adding to the first element $\lfloor \frac{r}{2^{p\alpha}} \rfloor$ integer multiples of $\frac{m}{2^{p\alpha}}$ so that the greatest element so-obtained is still less than m, provided that m is divisible by this power $2^{p\alpha}$. When m is no longer divisible by $2^{p\alpha}$, i.e. when $\alpha > \frac{z}{p}$, then we add

integer multiples of k, which is the odd part of m. In particular, if m is odd, i.e. if $z = 0$, then all the sets C'_α are reduced to a single state: $C'_\alpha = \{(\lfloor \frac{r}{2^{p\alpha}} \rfloor, T)\}$. Finally, note that since $R = |\text{rep}_{2^p}(r)|$ and $N \geqslant R$, we have $\lfloor \frac{r}{2^{pN}} \rfloor = 0$ and $C'_N = \{(k\ell, T_\ell) : 0 \leqslant \ell \leqslant 2^z - 1\}$.

We will see in Lemma 42 that the states in C'_α are exactly those from which there is a path labeled by the suffix of length α of $0^{N-R}\text{rep}_{2^p}(r)$ to the state (r, T).

Example 28. Let $m = 24$ and $p = 2$. We have $k = 3$ and $z = 3$. Let us consider the extremal possible values of the remainder r. For $r = 23$, we have $\text{rep}_4(23) = 113$, $R = 3$ and $N = \max\{\lceil \frac{3}{2} \rceil, 3\} = 3$. Thus, the sets defined above are

$$C'_0 = \{(23, T)\}$$
$$C'_1 = \{(5, T), (11, B), (17, B), (23, T)\}$$
$$C'_2 = \{(1, T), (4, B), (7, B), (10, T), (13, B), (16, T), (19, T), (22, B)\}$$
$$C'_3 = \{(0, T), (3, B), (6, B), (9, T), (12, B), (15, T), (18, T), (21, B)\}.$$

For instance, we have the following paths, which are labeled by $0^{N-R}\text{rep}_4(23) = 113$:

$$(0, T) \xrightarrow{1} (1, T) \xrightarrow{1} (5, T) \xrightarrow{3} (23, T)$$

and

$$(3, B) \xrightarrow{1} (13, B) \xrightarrow{1} (5, T) \xrightarrow{3} (23, T).$$

For $r = 0$, we have $R = 0$ and $N = \max\{\lceil \frac{3}{2} \rceil, 0\} = 2$. In this case, the sets C'_α are

$$C'_0 = \{(0, T)\}$$
$$C'_1 = \{(0, T), (6, B), (12, B), (18, T)\}$$
$$C'_2 = \{(0, T), (3, B), (6, B), (9, T), (12, B), (15, T), (18, T), (21, B)\}.$$

For instance, we have the following paths, which are labeled by $0^{N-R}\text{rep}_4(0) = 00$:

$$(0, T) \xrightarrow{0} (0, T) \xrightarrow{0} (0, T)$$

and

$$(3, B) \xrightarrow{0} (12, B) \xrightarrow{0} (0, T).$$

The sets C'_α are not necessarily disjoint as Example 28 shows. In order to obtain the desired classes of states of the automaton $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$, we consider the following definition.

Definition 29. For $\alpha \in [\![0, N]\!]$, we define

$$C_\alpha = C'_\alpha \setminus \bigcup_{\beta=0}^{\alpha-1} C'_\beta.$$

Let us define a second type of classes. The idea behind this definition is that these classes are "too far" from the remainder r with respect to the division by consecutive powers of the base 2^p, in the sense that these states do not accept any suffix of $0^{N-R}\text{rep}_{2^p}(r)$.

Definition 30. For $(j, X) \in \big([\![0, k-1]\!] \times \{T, B\}\big) \setminus \{(0, T)\}$, we define

$$D'_{(j,X)} = \{(j + \ell k, X_\ell) \colon 0 \leqslant \ell \leqslant 2^z - 1\}$$

and

$$D_{(j,X)} = D'_{(j,X)} \setminus \bigcup_{\alpha=0}^{N} C_\alpha.$$

As we already observed, all states of the form $(\ell k, T_\ell)$ appear in the set C'_N and thus, also in the union of the sets C_α. This is the reason why the sets $D'_{(0,T)}$ and $D_{(0,T)}$ are not defined, i.e. $(j, X) \neq (0, T)$ in the previous definition.

We will refer to the sets of states C_α and $D_{(j,X)}$ as *classes* of states. Let us make some preliminary observations concerning the previous definitions.

The classes C_α and $D_{(j,X)}$ are pairwise disjoint: the intersection of any two such classes is empty. Moreover, the nonempty classes C_α and $D_{(j,X)}$ form a partition of the whole set of states $[\![0, m-1]\!] \times \{T, B\}$ of $\Pi\left(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}\right)$. Note that if m is odd, i.e. if $z = 0$, then the sets $D'_{(j,X)}$ are reduced to a single state. If m is a power of 2, i.e. if $k = 1$, then there is no set of the form $D'_{(j,X)}$ and $D_{(j,X)}$ with $j \geqslant 1$.

Example 31. Let us resume Example 28. For $r = 23$, the classes defined above are

$$
\begin{aligned}
C_0 &= \{(23, T)\} \\
C_1 &= \{(5, T), (11, B), (17, B)\} \\
C_2 &= \{(1, T), (4, B), (7, B), (10, T), (13, B), (16, T), (19, T), (22, B)\} \\
C_3 &= \{(0, T), (3, B), (6, B), (9, T), (12, B), (15, T), (18, T), (21, B)\} \\
D_{(1,T)} &= \varnothing \\
D_{(2,T)} &= \{(2, T), (5, B), (8, B), (11, T), (14, B), (17, T), (20, T), (23, B)\} \\
D_{(0,B)} &= \{(0, B), (3, T), (6, T), (9, B), (12, T), (15, B), (18, B), (21, T)\} \\
D_{(1,B)} &= \{(1, B), (4, T), (7, T), (10, B), (13, T), (16, B), (19, B), (22, T)\} \\
D_{(2,B)} &= \{(2, B), (8, T), (14, T), (20, B)\}.
\end{aligned}
$$

Note that $2k + \lceil \frac{z}{p} \rceil = 2 \cdot 3 + \lceil \frac{3}{2} \rceil = 8$ of them are nonempty. In Figure 8, the automaton $\Pi\left(\mathcal{A}_{24,23,4} \times \mathcal{A}_{\mathcal{T},4}\right)$ is represented without the transitions and the states are colored with respect to these classes. If now we consider $r = 0$, these classes are

Figure 8: The classes of the projected automaton $\Pi\left(\mathcal{A}_{24,23,4} \times \mathcal{A}_{\mathcal{T},4}\right)$.

$$C_0 = \{(0, T)\}$$
$$C_1 = \{(6, B), (12, B), (18, T)\}$$
$$C_2 = \{(3, B), (9, T), (15, T), (21, B)\}$$
$$D_{(1,T)} = \{(1, T), (4, B), (7, B), (10, T), (13, B), (16, T), (19, T), (22, B)\}$$
$$D_{(2,T)} = \{(2, T), (5, B), (8, B), (11, T), (14, B), (17, T), (20, T), (23, B)\}$$
$$D_{(0,B)} = \{(0, B), (3, T), (6, T), (9, B), (12, T), (15, B), (18, B), (21, T)\}$$
$$D_{(1,B)} = \{(1, B), (4, T), (7, T), (10, B), (13, T), (16, B), (19, B), (22, T)\}$$
$$D_{(2,B)} = \{(2, B), (5, T), (8, T), (11, B), (14, T), (17, B), (20, B), (23, T)\}.$$

In this case, they are all nonempty and there are 8 of them. In Figure 9, the states of the automaton $\Pi\left(\mathcal{A}_{24,0,4} \times \mathcal{A}_{\mathcal{T},4}\right)$ are colored with respect to these classes.

Figure 9: The classes of the projected automaton $\Pi\left(\mathcal{A}_{24,0,4} \times \mathcal{A}_{\mathcal{T},4}\right)$.

Our aim is to prove that the nonempty classes defined above correspond exactly to the left quotients of the language $\mathrm{val}_{2^p}^{-1}(m\mathcal{T} + r)$.

6.2 Search for the empty classes

Proposition 32. *For all $\alpha \in [\![0, N]\!]$, the classes C_α are nonempty.*

Proof. The sequence $\left(\left\lfloor \frac{r}{2^{p\alpha}} \right\rfloor\right)_{\alpha \in [\![0, N]\!]}$ is (strictly) decreasing for $\alpha \in [\![0, R]\!]$ and is equal to 0 for $\alpha \in [\![R, N]\!]$. In particular, note that $\alpha = R$ is the first value for which $\left\lfloor \frac{r}{2^{p\alpha}} \right\rfloor = 0$. Therefore $(\left\lfloor \frac{r}{2^{p\alpha}} \right\rfloor, T_0) = (\left\lfloor \frac{r}{2^{p\alpha}} \right\rfloor, T) \in C_\alpha$ for every $\alpha \in [\![0, R]\!]$; see Figure 10. If $N = R$

C_0'	(r, T)	
C_1'	$(\left\lfloor \frac{r}{2^p} \right\rfloor, T)$	
$\vdots$	$\vdots$	
C_R'	$(0, T)$	
C_{R+1}'	$(0, T)$	$\left(\frac{m}{2^{p(R+1)}}, B\right)$
C_{R+2}'	$(0, T)$	$\left(\frac{m}{2^{p(R+2)}}, B\right)$
$\vdots$	$\vdots$	
C_N'	$(0, T)$	$\left(\frac{m}{2^{pN}}, B\right)$

Figure 10: The elements of the sets C_α' up to the first belonging to the classes C_α.

then we are done (in this case, the part of Figure 10 below the line is empty).

Now suppose that $N = \lceil \frac{z}{p} \rceil > R$. Then $(\lfloor \frac{r}{2^{p\alpha}} \rfloor, T_0) = (0, T) \in C'_\alpha \setminus C_\alpha$ for every $\alpha \in [\![R+1, N]\!]$. We claim that, for each such α, the "next" element of C'_α (i.e. the element corresponding to $\ell = 1$) indeed belongs to C_α. Let us fix some $\alpha \in [\![R+1, N]\!]$.

First, suppose that $p\alpha \leqslant z$. Then we have to prove that $(\frac{m}{2^{p\alpha}}, T_1) = (\frac{m}{2^{p\alpha}}, B) \notin C'_\beta$ for $\beta < \alpha$. If $\alpha > \beta$ then $\frac{m}{2^{p\alpha}} < \frac{m}{2^{p\beta}} \leqslant \lfloor \frac{r}{2^{p\beta}} \rfloor + \frac{m}{2^{p\beta}}$. This shows that if the state $(\frac{m}{2^{p\alpha}}, B)$ belongs to some C'_β with $\beta < \alpha$, then its first component has to be $\lfloor \frac{r}{2^{p\beta}} \rfloor$. But since it has second component $T_1 = B$ and since $T_0 = T$, the state $(\frac{m}{2^{p\alpha}}, B)$ cannot be the first state of any set C'_β.

Second, suppose that $p\alpha > z$. In this case, we have to prove that $(k, B) \notin C'_\beta$ for $\beta < \alpha$. Similarly to what precedes, if (k, B) belongs to some set C'_β, then we must have $\lfloor \frac{r}{2^{p\beta}} \rfloor = 0$ and $\beta \geqslant \frac{z}{p}$. But since $\frac{z}{p} < \alpha \leqslant N = \lceil \frac{z}{p} \rceil$, we get $\alpha = \lceil \frac{z}{p} \rceil$. Therefore, any $\beta < \alpha$ is such that $p\beta < z$. $\qquad\square$

Lemma 33. *We have* $|\mathrm{rep}_{2^p}(m-1)| - \lceil \frac{z}{p} \rceil \in [\![0, k-1]\!]$.

Proof. Observe that

$$|\mathrm{rep}_{2^p}(m)| = \lfloor \log_{2^p}(m) \rfloor + 1 = \left\lfloor \log_{2^p}(k) + \frac{z}{p} \right\rfloor + 1. \tag{2}$$

If m is a power of 2^p, i.e. if $k = 1$ and p divides z, then $|\mathrm{rep}_{2^p}(m-1)| = |\mathrm{rep}_{2^p}(m)| - 1 = \frac{z}{p}$ and the result is clear.

Now, suppose that m is not a power of the base 2^p. Then $|\mathrm{rep}_{2^p}(m-1)| = |\mathrm{rep}_{2^p}(m)|$. From (2) we get that $|\mathrm{rep}_{2^p}(m)| \geqslant \lceil \frac{z}{p} \rceil$. Let us show that $|\mathrm{rep}_{2^p}(m)| - \lceil \frac{z}{p} \rceil \leqslant k - 1$. If $k = 1$ then p does not divide z (for otherwise m would be a power of 2^p) and we get from (2) that $|\mathrm{rep}_{2^p}(m)| = \lfloor \frac{z}{p} \rfloor + 1 = \lceil \frac{z}{p} \rceil$. If $k = 3$ and $p = 1$, then we obtain $|\mathrm{rep}_{2^p}(m)| = \lfloor \log_2(3) + z \rfloor + 1 = 2 + z = k - 1 + \lceil \frac{z}{p} \rceil$. In all other cases, that is if $k \geqslant 5$ or ($k = 3$ and $p \geqslant 2$), we can check that $\log_{2^p}(k) < k - 2$. Therefore we have $|\mathrm{rep}_{2^p}(m)| \leqslant \lfloor k - 2 + \frac{z}{p} \rfloor + 1 = k - 1 + \lfloor \frac{z}{p} \rfloor$. $\qquad\square$

Proposition 34. *The empty classes* $D_{(j,X)}$ *are exactly those of the form* $D_{(\lfloor \frac{r}{2^{p\alpha}} \rfloor, T)}$ *with* $p\alpha \geqslant z$.

Proof. The k classes $D_{(0,B)}, \ldots, D_{(k-1,B)}$ are all nonempty since for any $j \in [\![0, k-1]\!]$, the state (j, B) does not belong to any C_α.

Now, let $j \in [\![1, k-1]\!]$. We have to show that all classes of the form $D_{(\lfloor \frac{r}{2^{p\alpha}} \rfloor, T)}$ with $p\alpha \geqslant z$ are empty (observe that if $p\alpha \geqslant z$ then $\lfloor \frac{r}{2^{p\alpha}} \rfloor \leqslant k - 1$), and that the other classes $D_{(j,T)}$ are nonempty.

If $(j, T) \notin \cup_{\alpha=0}^N C'_\alpha$, then the class $D_{(j,T)}$ is nonempty since it contains (j, T). In this case, $j \neq \lfloor \frac{r}{2^{p\alpha}} \rfloor$ for any $\alpha \in [\![0, N]\!]$. Now, suppose that there exists some $\alpha \in [\![0, N]\!]$ such that $(j, T) \in C'_\alpha$. Since $j < k$, this α is unique and $j = \lfloor \frac{r}{2^{p\alpha}} \rfloor$. We have to show that $D_{(j,T)}$ is empty if and only if $p\alpha \geqslant z$.

Clearly, $p\alpha \geqslant z$ implies that $D'_{(j,T)} = \{(j + \ell k, T_\ell): \ell \in [\![0, 2^z - 1]\!]\} = C'_\alpha$, and hence that $D_{(j,T)}$ is empty.

Now suppose that $p\alpha < z$. We show that the second element $(j + k, B)$ of the set $D'_{(j,T)}$ does not belong to any set C'_β, and hence indeed belongs to the class $D_{(j,T)}$. Let $\beta \in [\![0, N]\!]$ and suppose to the contrary that $(j + k, B) \in C'_\beta$. Since $j + k \in [\![0, 2k - 1]\!]$, the state $(j + k, B)$ must be either the first or the second element of the set C'_β. But since $B \neq T_0 = T$, it has to be the second. If $p\beta < z$, then we obtain $j + k = \lfloor \frac{r}{2^{p\beta}} \rfloor + k2^{z-p\beta} \geqslant 2k$, a contradiction. Thus $p\beta \geqslant z$ and $j + k = \lfloor \frac{r}{2^{p\beta}} \rfloor + k$. But this implies that $\lfloor \frac{r}{2^{p\alpha}} \rfloor = j = \lfloor \frac{r}{2^{p\beta}} \rfloor$. Since $\beta > \alpha$, this means that $j = 0$, a contradiction. $\qquad\square$

Corollary 35. *There are exactly $N - \lceil \frac{z}{p} \rceil$ empty classes among the $2k - 1$ classes $D_{(j,X)}$.*

Proof. By Proposition 34, the k classes $D_{(0,B)}, \ldots, D_{(k-1,B)}$ are nonempty and we have to count the number of classes of the form $D_{(\lfloor \frac{r}{2^{p\alpha}} \rfloor, T)}$ with $p\alpha \geqslant z$ among the $k - 1$ classes $D_{(1,T)}, \ldots, D_{(k-1,T)}$. Note that by Lemma 33 and by definition of N, we have $N - \lceil \frac{z}{p} \rceil \in [\![0, k - 1]\!]$.

Equivalently, we have to count the elements $\alpha \in [\![\lceil \frac{z}{p} \rceil, N]\!]$ such that $\lfloor \frac{r}{2^{p\alpha}} \rfloor \neq 0$. Similarly to the proof of Proposition 32, we consider two cases (also see Figure 10). If $N = \lceil \frac{z}{p} \rceil$ then there is no such α at all since $\lfloor \frac{r}{2^{pN}} \rfloor = 0$. If $N = R > \lceil \frac{z}{p} \rceil$, then the suitable α are exactly those in $[\![\lceil \frac{z}{p} \rceil, R - 1]\!]$, and there are exactly $R - 1 - \lceil \frac{z}{p} \rceil + 1 = N - \lceil \frac{z}{p} \rceil$ of them. Hence the conclusion. $\qquad\square$

6.3 States of the same class are indistinguishable

In order to prove that two states (j, X) and (j', X') of the automaton $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$ are indistinguishable, we have to prove that $L_{(j,X)} = L_{(j',X')}$. The general procedure that we use for proving that $L_{(j,X)} = L_{(j',X')}$ goes as follows. Pick some word $v \in A_{2^p}^*$ and let $n = |v|$ and $e = \mathrm{val}_{2^p}(v)$. By Lemma 12, the word v is accepted from the state (j, X) if and only if there exists some $d \in \mathbb{N}$ such that

$$2^{pn}j + e = md + r \quad\text{and}\quad X_d = T.$$

Similarly, the word v is accepted from the state (j', X') if and only if there exists some $d' \in \mathbb{N}$ such that

$$2^{pn}j' + e = md' + r \quad\text{and}\quad (X')_{d'} = T.$$

But then, observe that there is only one possible pair of candidates for d and d': we necessarily have

$$d = \frac{2^{pn}j + e - r}{m} \quad\text{and}\quad d' = \frac{2^{pn}j' + e - r}{m}. \tag{3}$$

Therefore, proving that

$$L_{(j,X)} = L_{(j',X')}$$

is equivalent to proving that for all $n \in \mathbb{N}$ and $e \in [\![0, 2^{pn} - 1]\!]$, we have

$$(d \in \mathbb{N} \text{ and } X_d = T) \iff (d' \in \mathbb{N} \text{ and } (X')_{d'} = T)$$

where d and d' are given by (3). Moreover, note such d and d' are always greater than or equal to $-\frac{r}{m}$, hence they are greater than -1. Thus, provided that d and d' are integers,

we know that they are necessary nonnegative. Similarly, thanks to Remark 13, d and d' must be less than 2^{pn}. For these reasons, in the forthcoming proofs (namely, in Lemmas 36 and 38), we need to verify that $d, d' \in \mathbb{Z}$ but we don't need to check that $0 \leqslant d, d' < 2^{pn}$.

Our first aim is to show that all states in the same class $D_{(j,X)}$ accept the same language. We start with a lemma that will be used several times. Note that this lemma does not only concern the classes $D_{(j,X)}$ since we can have $(j, X) = (0, T)$ in the statement.

Lemma 36. *Let* $j \in [\![0, k-1]\!]$, $\ell \in [\![0, 2^z - 1]\!]$ *and* $X \in \{T, B\}$. *For all* $n \in \mathbb{N}$ *such that* $pn \geqslant z$, *we have*
$$L_{(j,X)} \cap (A_{2^p})^n = L_{(j+\ell k, X_\ell)} \cap (A_{2^p})^n.$$

Proof. Let $n \in \mathbb{N}$ such that $pn \geqslant z$ and let $e \in [\![0, 2^{pn} - 1]\!]$. Set
$$d = \frac{2^{pn}j + e - r}{m} \quad \text{and} \quad d' = \frac{2^{pn}(j + \ell k) + e - r}{m}.$$

Following the procedure described above, we have to prove that $(d \in \mathbb{N}$ and $X_d = T) \iff (d' \in \mathbb{N}$ and $(X_\ell)_{d'} = T)$. Since $d' = d + \frac{2^{pn}\ell k}{m} = d + \ell 2^{pn-z}$ and since $pn \geqslant z$, d is an integer if and only if so is d'. Moreover,
$$d \leqslant \frac{2^{pn}j + e}{m} < \frac{2^{pn}(j+1)}{m} \leqslant \frac{2^{pn}k}{m} = 2^{pn-z}. \tag{4}$$

If $d, d' \in \mathbb{N}$ then $\mathrm{rep}_2(d') = \mathrm{rep}_2(\ell)0^{pn-z-|\mathrm{rep}_2(d)|}\mathrm{rep}_2(d)$, and hence $X_d = (X_\ell)_{d'}$. $\square$

Proposition 37. *Let* $(j, X) \in ([\![0, k-1]\!] \times \{T, B\}) \setminus \{(0, T)\}$. *Then any two states in* $D_{(j,X)}$ *accept the same language.*

Proof. Let $\ell, \ell' \in [\![0, 2^z - 1]\!]$. It suffices to show that if $(j + \ell k, X_\ell) \in D_{(j,X)}$ then $L_{(j+\ell k, X_\ell)} \subseteq L_{(j+\ell' k, X_{\ell'})}$. Thus, suppose that $(j+\ell k, X_\ell) \notin \cup_{\alpha=0}^{N} C_\alpha$. Let $n \in \mathbb{N}$ and $e \in [\![0, 2^{pn} - 1]\!]$. Set $d = \frac{2^{pn}(j+\ell k)+e-r}{m}$ and assume that $d \in \mathbb{N}$ and $(X_\ell)_d = T$. Then $X_\ell = T_d$. If $pn < z$ then $\frac{r-e}{2^{pn}} = \lfloor \frac{r}{2^{pn}} \rfloor$ because $\frac{r-e+dm}{2^{pn}} = j+k\ell$ is an integer, m is divisible by 2^{pn} and $e \in [\![0, 2^{pn}-1]\!]$. Therefore, if $pn < z$ then we get that
$$(j + \ell k, X_\ell) = \left(\frac{r - e + dm}{2^{pn}}, T_d \right) = \left(\left\lfloor \frac{r}{2^{pn}} \right\rfloor + d\frac{m}{2^{pn}}, T_d \right) \in C'_n$$

which contradicts our assumption. So $pn \geqslant z$ and the conclusion follows from Lemma 36. $\square$

Note that the proof of Proposition 37 shows that no word shorter than $\lfloor \frac{z}{p} \rfloor$ is accepted from a state of a class $D_{(j,X)}$. However, such words may be accepted from a state of one of the classes C_α (see Lemma 42 below).

Now we turn to the classes C_α. The proof is divided in several technical lemmas.

Lemma 38. *For every* $\alpha \in [\![0, N]\!]$, *any two states in* C'_α *accept the same words of length at least* α.

Proof. Let $\alpha \in [\![0, N]\!]$. First, we do the case $\alpha \leqslant \frac{z}{p}$. By definition of the sets C'_α, it suffices to show that for all $\ell \in [\![0, 2^{p\alpha}-1]\!]$ and $n \geqslant \alpha$, we have $L_{(\lfloor \frac{r}{2^{p\alpha}} \rfloor, T)} \cap (A_{2^p})^n = L_{(\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell \frac{m}{2^{p\alpha}}, T_\ell)} \cap (A_{2^p})^n$. Thus, let $\ell \in [\![0, 2^{p\alpha}-1]\!]$, $n \geqslant \alpha$ and $e \in [\![0, 2^{pn}-1]\!]$. Then set

$$d = \frac{2^{pn}\lfloor \frac{r}{2^{p\alpha}} \rfloor + e - r}{m} \quad \text{and} \quad d' = \frac{2^{pn}(\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell \frac{m}{2^{p\alpha}}) + e - r}{m}.$$

We have to prove that $(d \in \mathbb{N} \text{ and } T_d = T) \iff (d' \in \mathbb{N} \text{ and } (T_\ell)_{d'} = T)$. Since $\lfloor \frac{r}{2^{p\alpha}} \rfloor < \frac{m}{2^{p\alpha}} = k2^{z-p\alpha}$ and $z - p\alpha \geqslant 0$, we obtain that $\lfloor \frac{r}{2^{p\alpha}} \rfloor + 1 \leqslant \frac{m}{2^{p\alpha}}$. Then

$$d < \frac{2^{pn}\left(\lfloor \frac{r}{2^{p\alpha}} \rfloor + 1\right)}{m} \leqslant 2^{p(n-\alpha)}. \tag{5}$$

Since $d' = d + \ell 2^{p(n-\alpha)}$ and $n \geqslant \alpha$, it follows that d is an integer if and only if so is d'. Moreover, if both d and d' are in $\mathbb{N}$ then $\mathrm{rep}_2(d') = \mathrm{rep}_2(\ell)0^{p(n-\alpha)-|\mathrm{rep}_2(d)|}\mathrm{rep}_2(d)$, hence $T_d = (T_\ell)_{d'}$.

Second, suppose that $\alpha > \frac{z}{p}$. In this case, we have to show that for all $\ell \in [\![0, 2^z-1]\!]$ and $n \geqslant \alpha$, we have $L_{(\lfloor \frac{r}{2^{p\alpha}} \rfloor, T)} \cap (A_{2^p})^n = L_{(\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell k, T_\ell)} \cap (A_{2^p})^n$. Since $\lfloor \frac{r}{2^{p\alpha}} \rfloor < \frac{m}{2^{p\alpha}} = k2^{z-p\alpha} < k$, the conclusion follows from Lemma 36. $\qquad \square$

Lemma 39. *Let $\alpha \in [\![0, N]\!]$.*

1. If $p\alpha \leqslant z$ then no state in C_α accepts any words of length $< \alpha$.

2. If $p\alpha > z$ then no state in C_α accepts any words of length $\leqslant \lfloor \frac{z}{p} \rfloor$.

Proof. Let us prove item 1. Suppose that $p\alpha \leqslant z$ and that there exists a word over A_{2^p} of length $\beta < \alpha$ that is accepted from a state of the form $(\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell \frac{m}{2^{p\alpha}}, T_\ell)$ with $\ell \in [\![0, 2^{p\alpha} - 1]\!]$, i.e. from a state in C'_α. This means that there exists $e \in [\![0, 2^{p\beta} - 1]\!]$ such that if we set

$$d = \frac{2^{p\beta}(\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell \frac{m}{2^{p\alpha}}) + e - r}{m},$$

then $d \in \mathbb{N}$ and $(T_\ell)_d = T$. But then

$$\left(\left\lfloor \frac{r}{2^{p\alpha}} \right\rfloor + \ell \frac{m}{2^{p\alpha}}, T_\ell\right) = \left(\frac{r - e + dm}{2^{p\beta}}, T_d\right) = \left(\left\lfloor \frac{r}{2^{p\beta}} \right\rfloor + d\frac{m}{2^{p\beta}}, T_d\right) \in C'_\beta.$$

Since $\beta < \alpha$, the state $(\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell \frac{m}{2^{p\alpha}}, T_\ell)$ does not belong to C_α.

Now we prove item 2. Suppose that $p\alpha > z$ and that there exists a word over A_{2^p} of length $\beta \leqslant \lfloor \frac{z}{p} \rfloor$ that is accepted from a state of the form $(\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell k, T_\ell)$ with $\ell \in [\![0, 2^z-1]\!]$, i.e. from a state in C'_α. This means that there exists $e \in [\![0, 2^{p\beta} - 1]\!]$ such that if we set

$$d = \frac{2^{p\beta}(\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell k) + e - r}{m},$$

then $d \in \mathbb{N}$ and $(T_\ell)_d = T$. But then

$$\left(\left\lfloor \frac{r}{2^{p\alpha}} \right\rfloor + \ell k, T_\ell\right) = \left(\frac{r - e + dm}{2^{p\beta}}, T_d\right) = \left(\left\lfloor \frac{r}{2^{p\beta}} \right\rfloor + d\frac{m}{2^{p\beta}}, T_d\right) \in C'_\beta.$$

Therefore the state $(\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell k, T_\ell)$ does not belong to C_α. $\qquad \square$

Lemma 40. *If $N = \lceil \frac{z}{p} \rceil$, then no state in C_N accepts any words of length $< N$.*

Proof. Suppose that $N = \lceil \frac{z}{p} \rceil$ and that v is a word over A_{2^p} of length $\beta < N$ that is accepted from a state of the form $(\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell k, T_\ell)$ with $\ell \in [\![0, 2^z - 1]\!]$, i.e. from a state in C'_N. Set

$$e = \mathrm{val}_{2^p}(v) \quad \text{and} \quad d = \frac{2^{p\beta}(\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell k) + e - r}{m}.$$

Then $d \in \mathbb{N}$ and $(T_\ell)_d = T$. We get

$$\left\lfloor \frac{r}{2^{p\alpha}} \right\rfloor + \ell k = \frac{r - e + dm}{2^{p\beta}} = \left\lfloor \frac{r}{2^{p\beta}} \right\rfloor + d \frac{m}{2^{p\beta}}$$

and $T_\ell = T_d$. Since $\beta < \frac{z}{p}$, this shows that the state $(\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell k, T_\ell) = (\lfloor \frac{r}{2^{p\beta}} \rfloor + d \frac{m}{2^{p\beta}}, T_d)$ belongs to C'_β, and hence cannot belong to C_N. $\square$

We are now ready to prove that two states belonging to any given class C_α are indistinguishable.

Proposition 41. *For every $\alpha \in [\![0, N]\!]$, any two states in C_α accept the same language.*

Proof. Let $\alpha \in [\![0, N]\!]$. From Lemma 38, it is enough to consider words of length smaller than α and from the first item of Lemma 39, we may suppose that $p\alpha > z$. If $N = \lceil \frac{z}{p} \rceil$, then we must have $\alpha = N$ and we are done thanks to Lemma 40. Thus, we may also assume that $N = R > \lceil \frac{z}{p} \rceil$. Under these assumptions, $\lfloor \frac{r}{2^{p\alpha}} \rfloor < k$ and the first state of C_α is $(\lfloor \frac{r}{2^{p\alpha}} \rfloor, T)$; see Figure 10. Thus, we have to show that for all $\ell \in [\![0, 2^z - 1]\!]$ such that the state $(\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell k, T_\ell)$ indeed belongs to C_α and all $n < \alpha$, we have

$$L_{(\lfloor \frac{r}{2^{p\alpha}} \rfloor, T)} \cap (A_{2^p})^n = L_{(\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell k, T_\ell)} \cap (A_{2^p})^n.$$

If $pn < z$ then both languages are empty by the second item of Lemma 39. If $pn \geqslant z$ then the equality follows from Lemma 36. $\square$

6.4 States of different classes are distinguishable

In this section, we show that, in the projected automaton $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$, states belonging to different classes C_α or $D_{(j,X)}$ are pairwise distinguishable, that is, for any two such states, there exists a word which is accepted from exactly one of them.

The following lemma shows that the states in a set C'_α are exactly those states that leads to states in $C'_{\alpha-1}$ by reading the letter $r_{\alpha-1}$, where $0^{N-R}\mathrm{rep}_{2^p}(r) = r_{N-1} \cdots r_1 r_0$.

Lemma 42. *Let $0^{N-R}\mathrm{rep}_{2^p}(r) = r_{N-1} \cdots r_1 r_0$ and let $\alpha \in [\![0, N]\!]$. Then*

$$C'_\alpha = \{(i, X) \in [\![0, m-1]\!] \times \{T, B\} : \delta^\Pi_\times((i, X), r_{\alpha-1} \cdots r_1 r_0) = (r, T)\}.$$

Proof. First, we consider the case where $p\alpha \leqslant z$. Pick some $(i, X) \in C'_\alpha$. By definition, there exists $\ell \in [\![0, 2^{p\alpha} - 1]\!]$ such that $(i, X) = (\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell\frac{m}{2^{p\alpha}}, T_\ell)$. Observe that $\lfloor \frac{r}{2^{p\alpha}} \rfloor = \mathrm{val}_{2^p}(r_{N-1} \cdots r_{\alpha+1} r_\alpha)$. Then

$$2^{p\alpha}\left(\left\lfloor \frac{r}{2^{p\alpha}} \right\rfloor + \ell\frac{m}{2^{p\alpha}}\right) + \mathrm{val}_{2^p}(r_{\alpha-1} \cdots r_1 r_0) = \ell m + r.$$

Since $(T_\ell)_\ell = T$, we obtain from Lemma 12 that $\delta_\times^\Pi((i, X), r_{\alpha-1} \cdots r_1 r_0) = (r, T)$. Conversely, pick some state $(i, X) \in [\![0, m-1]\!] \times \{T, B\}$ such that $\delta_\times^\Pi((i, X), r_{\alpha-1} \cdots r_1 r_0) = (r, T)$. Then there exists some $d \in [\![0, 2^{p\alpha} - 1]\!]$ such that

$$2^{p\alpha}i + \mathrm{val}_{2^p}(r_{\alpha-1} \cdots r_1 r_0) = md + r \quad \text{and} \quad X_d = T.$$

We obtain

$$i = \frac{1}{2^{p\alpha}}\left(md + r - \mathrm{val}_{2^p}(r_{\alpha-1} \cdots r_1 r_0)\right) = d\frac{m}{2^{p\alpha}} + \left\lfloor \frac{r}{2^{p\alpha}} \right\rfloor.$$

Observe that $X_d = T$ is equivalent to $X = T_d$. This proves that $(i, X) \in C'_\alpha$.

Second, we consider the case where $p\alpha > z$. Pick some $(i, X) \in C'_\alpha$. There exists $\ell \in [\![0, 2^z - 1]\!]$ such that $(i, X) = (\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell k, T_\ell)$. Then

$$2^{p\alpha}\left(\left\lfloor \frac{r}{2^{p\alpha}} \right\rfloor + \ell k\right) + \mathrm{val}_{2^p}(r_{\alpha-1} \cdots r_1 r_0) = \ell k 2^{p\alpha} + r = \ell 2^{p\alpha - z} m + r.$$

Since $(T_\ell)_{\ell 2^{p\alpha-z}} = (T_\ell)_\ell = T$, we obtain that $\delta_\times^\Pi((i, X), r_{\alpha-1} \cdots r_1 r_0) = (r, T)$. Conversely, pick some state $(i, X) \in [\![0, m-1]\!] \times \{T, B\}$ such that $\delta_\times^\Pi((i, X), r_{\alpha-1} \cdots r_1 r_0) = (r, T)$. Then there exists some $d \in [\![0, 2^{p\alpha} - 1]\!]$ such that

$$2^{p\alpha}i + \mathrm{val}_{2^p}(r_{\alpha-1} \cdots r_1 r_0) = md + r \quad \text{and} \quad X_d = T.$$

From the first equality, we have $md = 2^{p\alpha}i - (r - \mathrm{val}_{2^p}(r_{\alpha-1} \cdots r_1 r_0)) = 2^{p\alpha}i - 2^{p\alpha}\lfloor \frac{r}{2^{p\alpha}} \rfloor$, hence $kd = 2^{p\alpha-z}(i - \lfloor \frac{r}{2^{p\alpha}} \rfloor)$. Since k is odd, d must be a multiple of $2^{p\alpha-z}$. We obtain

$$i = \frac{d}{2^{p\alpha-z}}k + \left\lfloor \frac{r}{2^{p\alpha}} \right\rfloor$$

and $X_{\frac{d}{2^{p\alpha-z}}} = X_d = T$. Since $\frac{d}{2^{p\alpha-z}} \in [\![0, 2^z - 1]\!]$, we get that $(i, X) \in C'_\alpha$. $\qquad\square$

Proposition 43. *For every $\alpha \in [\![0, N]\!]$, the class C_α is distinguishable from all the other classes.*

Proof. First, we show that the classes C_α are distinguishable among themselves. From Proposition 32, we know that these classes are all nonempty. Let $\alpha, \beta \in [\![0, N]\!]$ such that $\alpha < \beta$ and let $(i, X) \in C_\alpha$ and $(j, Y) \in C_\beta$. We show that $L_{(i,X)} \neq L_{(j,Y)}$. By definition of the classes, the state (i, X) belongs to C'_α and since $\alpha < \beta$, the state (j, Y) does not belong to C'_α. We get from Lemma 42 that the suffix s of length α of the word $0^{N-R}\mathrm{rep}_{2^p}(r)$ is accepted from (i, X) but not from (j, Y). So $s \in L_{(i,X)} \setminus L_{(j,Y)}$.

Second, we show that the classes C_α are distinguishable from all the nonempty classes of the form $D_{(i,X)}$. Let $\alpha \in [\![0, N]\!]$. By definition, any state in a class $D_{(i,X)}$ cannot belong to C'_α. Similarly to what precedes, the conclusion follows from Lemma 42. $\qquad\square$

It remains to show that the nonempty classes $D_{(j,X)}$ are distinguishable from each other.

Proposition 44. *Suppose that $k > 1$ and let $(i, X), (j, Y) \in (\llbracket 0, k-1 \rrbracket \times \{T, B\}) \setminus \{(0, T)\}$ be distinct and such that the classes $D_{(i,X)}$ and $D_{(j,Y)}$ are both nonempty. Then $D_{(i,X)}$ and $D_{(j,Y)}$ are distinguishable.*

Proof. We already know from the previous section that the states of $D_{(i,X)}$ (resp. $D_{(j,Y)}$) are indistinguishable. Therefore, it suffices to show that $L_{(i,X)} \neq L_{(j,Y)}$.

First, suppose that $i = j$. Then $X \neq Y$ by hypothesis and the states (i, X) and (j, Y) are disjoint by Lemma 25. Since $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$ is coaccessible by Proposition 26, we obtain that the states (i, X) and (j, Y) are distinguishable.

Now suppose that $i \neq j$. By Lemma 18, the word $w_i \mathrm{rep}_{2^p}(r)$ is accepted from i in the automaton $\Pi(\mathcal{A}_{m,r,2^p})$ but is not accepted from j. Then, there exist a word u_1 of length $|w_i|$ and a word u_2 of length R such that the word $(u_1, w_i)(u_2, \mathrm{rep}_{2^p}(r))$ is accepted from i in the automaton $\mathcal{A}_{m,r,2^p}$ but is not accepted from j. By Lemma 21, this word is accepted either from (i, T) or from (i, B) in the automaton $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$ but is not accepted neither from (j, T) nor from (j, B). Now, two cases are possible.

First, suppose that $(u_1, w_i)(u_2, \mathrm{rep}_{2^p}(r))$ is accepted from (i, X) in $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$. Then, in the projection $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$, the word $w_i \mathrm{rep}_{2^p}(r)$ is accepted from (i, X) but not from (j, Y). Thus, the word $w_i \mathrm{rep}_{2^p}(r)$ distinguishes the states (i, X) and (j, Y).

Second, suppose that $(u_1, w_i)(u_2, \mathrm{rep}_{2^p}(r))$ is accepted from $(i, \overline{X})$ in $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$. Let $(i', X') = \delta_\times((i, \overline{X}), (u_1, w_i))$. Then $\delta_\times((i', X'), (u_2, \mathrm{rep}_{2^p}(r))) = (r, T)$. In particular $\delta_{m,r,2^p}^\Pi(i, w_i) = i'$, hence $i' = 0$ by Lemma 17. Now, by using Lemma 21 and Lemma 22 successively, we obtain

$$
\begin{aligned}
\delta_\times\big((i, X), (u_1, w_i)\mathrm{rep}_{2^p}(1, m)(u_2, \mathrm{rep}_{2^p}(r))\big) &= \delta_\times\big((0, \overline{X'}), \mathrm{rep}_{2^p}(1, m)(u_2, \mathrm{rep}_{2^p}(r))\big) \\
&= \delta_\times\big((0, X'), (u_2, \mathrm{rep}_{2^p}(r))\big) \\
&= (r, T).
\end{aligned}
$$

This shows that the word $w_i \mathrm{rep}_{2^p}(m)\mathrm{rep}_{2^p}(r)$ is accepted from (i, X) in $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$. From Lemmas 18 and 24, this word cannot be accepted from (j, Y), hence it distinguishes the states (i, X) and (j, Y). $\qquad\square$

6.5 The minimal automaton of $\mathrm{val}_{2^p}^{-1}(m\mathcal{T} + r)$.

We are ready to construct the minimal automaton of $\mathrm{val}_{2^p}^{-1}(m\mathcal{T} + r)$. Since the states of $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$ that belong to the same class C_α or $D_{(j,X)}$ are indistinguishable, they can be glued together in order to define a new automaton $\mathcal{M}_{m,r,\mathcal{T},2^p}$ that still accepts the same language.

The formal definition of $\mathcal{M}_{m,r,\mathcal{T},2^p}$ is as follows. The alphabet is A_{2^p}. The states are the classes C_α for $\alpha \in \llbracket 0, N \rrbracket$ and the nonempty classes $D_{(j,X)}$ for $(j, X) \in (\llbracket 0, k-1 \rrbracket \times \{T, B\}) \setminus \{(0, T)\}$. The class C_R is the initial state and the only final state is the class C_0. Note that $(0, T) \in C_R$ and that $(r, T) \in C_0$. The transitions of $\mathcal{M}_{m,r,\mathcal{T},2^p}$ are defined

as follows: there is a transition labeled by a letter a in A_{2^p} from a class J_1 to a class J_2 if and only if in the automaton $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$, there is a transition labeled by a from a state of J_1 to a state of J_2.

Example 45. In Figure 11, the classes of $\Pi(\mathcal{A}_{6,2,4} \times \mathcal{A}_{\mathcal{T},4})$ are colored in white, blue, grey, yellow, fuchsia, orange and purple.
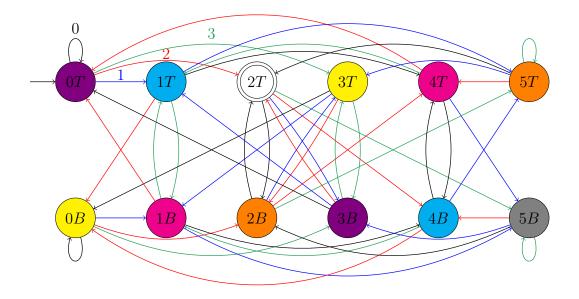

Figure 11: The classes of the automaton of $\Pi(\mathcal{A}_{6,2,4} \times \mathcal{A}_{\mathcal{T},4})$.

Figure 12 depicts the minimal automaton $\mathcal{M}_{6,2,\mathcal{T},4}$ of $\mathrm{val}_4^{-1}(6\mathcal{T}+2)$, where states corresponding to the same color are glued together to form a single state.

Theorem 46. *Let p and m be positive integers. The automaton $\mathcal{M}_{m,r,\mathcal{T},2^p}$ is the minimal automaton of the language $\mathrm{val}_{2^p}^{-1}(m\mathcal{T}+r)$.*

Proof. By construction and by Propositions 37 and 41, the language accepted by $\mathcal{M}_{m,r,\mathcal{T},2^p}$ is $\mathrm{val}_{2^p}^{-1}(m\mathcal{T}+r)$. In order to see that $\mathcal{M}_{m,r,\mathcal{T},2^p}$ is minimal, it suffices to prove that it is complete, reduced and accessible. The fact that $\mathcal{M}_{m,r,\mathcal{T},2^p}$ is reduced follows from Propositions 43 and 44. We know from Proposition 26 that the automaton $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$ is complete and accessible, which in turn implies that $\mathcal{M}_{m,r,\mathcal{T},2^p}$ is complete and accessible as well. $\qquad\square$

We are now ready to prove Theorem 3.

Proof of Theorem 3. In view of Theorem 46, it suffices to count the number of states of $\mathcal{M}_{m,r,\mathcal{T},2^p}$. These states correspond to the nonempty classes C_α and $D_{(j,X)}$. By Proposition 32 and Corollary 35, there are exactly $(N+1) + \left(2k-1-\left(N-\left\lceil\frac{z}{p}\right\rceil\right)\right) = 2k+\left\lceil\frac{z}{p}\right\rceil$ such classes. $\qquad\square$
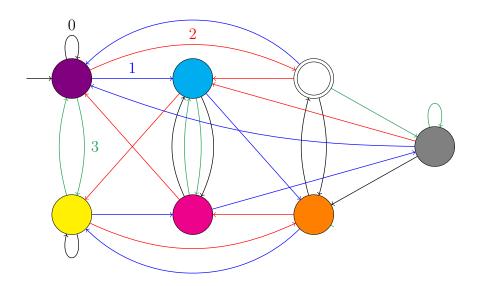
Figure 12: The automaton $\mathcal{M}_{6,2,\mathcal{T},4}$.

Example 47. The minimal automaton of the language $\mathrm{val}_4^{-1}(6\mathcal{T} + 2)$ has 7 states; see Figure 12. We can indeed compute that $2 \cdot 3 + \lceil \frac{1}{2} \rceil = 7$.

As already mentioned, the obtained state complexity of $m\mathcal{T} + r$ is independent of the considered remainder r. This observation is all the more surprising that the classes used in our construction indeed depend on r.

7 A direct description of the classes whenever $r = 0$

In the conference paper [14], we described the automaton $\mathcal{M}_{m,r,\mathcal{T},2^p}$ in the particular case where $r = 0$, i.e. for the exact multiples of $\mathcal{T}$. The construction was similar, but the way we built the classes of states was different. Therefore, we can give another description of the classes C_α and $D_{(j,X)}$ for $r = 0$ which is easier than the descriptions from Definitions 29 and 30 in the sense that the classes are built in a direct way, without having to remove some states a posteriori.

Note that if $r = 0$ then $R = 0$ and $N = \lceil \frac{z}{p} \rceil$.

Corollary 48. *Suppose that $r = 0$.*

- *We have $C_0 = \{(0, T)\}$.*

- *For each $\alpha \in [\![1, N{-}1]\!]$, we have*

$$C_\alpha = \bigcup_{\beta=(\alpha-1)p}^{\alpha p - 1} \{(k2^{z-\beta-1} + \ell k2^{z-\beta}, B_\ell) \colon \ell \in [\![0, 2^\beta{-}1]\!]\}.$$

- *We have*

$$C_N = \bigcup_{\beta=\left(\lceil \frac{z}{p} \rceil - 1\right)p}^{z-1} \{(k2^{z-\beta-1} + \ell k 2^{z-\beta}, B_\ell) \colon \ell \in [\![0, 2^\beta - 1]\!]\}.$$

- *For* $(j, X) \in \big([\![0, k-1]\!] \times \{T, B\}\big) \setminus \{(0, T)\}$, *we have*

$$D_{(j,X)} = \{(j + \ell k, X_\ell) \colon \ell \in [\![0, 2^z - 1]\!]\}.$$

Proof. This is a consequence of Theorem 46 and [14, Theorem 33]. $\qquad\square$

8 Replacing $\mathcal{T}$ by its complement $\overline{\mathcal{T}}$

If we are interested in the set $\overline{\mathcal{T}} = \mathbb{N} \setminus \mathcal{T}$ instead of $\mathcal{T}$, we can use the same construction that we described and studied for $\mathcal{T}$. We only have to exchange the final/non-final status of the states in the automaton $\mathcal{A}_\mathcal{T}$. In this section, we show that we may instead directly obtain the minimal automaton of the language $\mathrm{val}_{2^p}^{-1}(m\overline{\mathcal{T}} + r)$ from that of $\mathrm{val}_{2^p}^{-1}(m\mathcal{T} + r)$.

Example 49. Let us push further our running example by considering now $\overline{\mathcal{T}}$ instead of $\mathcal{T}$. The classes of states are defined similarly by exchanging T and B everywhere. In Figure 13, we have depicted the classes of the corresponding projected product automaton, which we denote by $\Pi\big(\mathcal{A}_{6,2,4} \times \mathcal{A}_{\overline{\mathcal{T}},4}\big)$. Figure 14 depicts the minimal automaton $\mathcal{M}_{6,2,\overline{\mathcal{T}},4}$
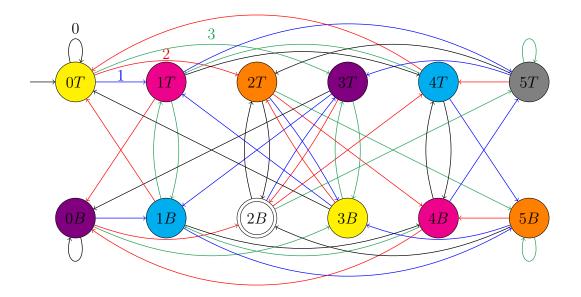

Figure 13: The classes of the automaton of $\Pi\left(\mathcal{A}_{6,2,4} \times \mathcal{A}_{\overline{\mathcal{T}},4}\right)$.

of $\mathrm{val}_4^{-1}(6\overline{\mathcal{T}} + 2)$, where states corresponding to the same color are glued together to form a single state. Since the classes of states have been modified but the edges are unchanged,
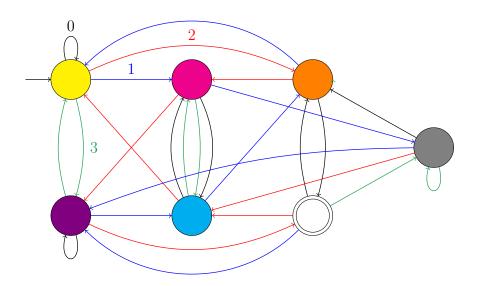
Figure 14: The automaton $\mathcal{M}_{6,2,\overline{\mathcal{T}},4}$.

the minimal automaton obtained by gluing the sets of the same classes together is not a symmetric version of the automaton $\mathcal{M}_{m,r,\mathcal{T},2^p}$ we obtained starting from the set $\mathcal{T}$; compare Figures 12 and 14. Nevertheless, observe that the automaton of Figure 14 can be obtained from the one of Figure 12 by replacing the initial state (in purple) by the yellow state. Also observe that, in the automaton of Figure 12, the yellow state is reached from the initial state by reading the word $\mathrm{rep}_4(6) = 12$. This fact is always true and is proved in Proposition 50.

In the next proposition, we show that the minimal automaton of $\mathrm{val}^{-1}(m\overline{\mathcal{T}} + r)$ can be obtained directly from the minimal automaton of $\mathrm{val}_{2^p}^{-1}(m\mathcal{T} + r)$ by only moving the initial state.

Proposition 50. *The minimal automaton of* $\mathrm{val}_{2^p}^{-1}(m\overline{\mathcal{T}} + r)$ *is obtained by replacing the initial state of the automaton* $\mathcal{M}_{m,r,\mathcal{T},2^p}$ *by the state that is reached by reading* $\mathrm{rep}_{2^p}(m)$ *from the initial state.*

Proof. Consider the automaton $\mathcal{M}_{m,r,\mathcal{T},2^p}$. By construction, its states are sets of states (called classes) of the automaton $\Pi\left(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}\right)$. By Lemma 12, for each $X \in \{T, B\}$, there is a path labeled by $\mathrm{rep}_{2^p}(m)$ going from $(0, X)$ to $(0, \overline{X})$ in $\Pi\left(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}\right)$, and hence the same holds for the corresponding classes of states in $\mathcal{M}_{m,r,\mathcal{T},2^p}$.

First, let us show that the obtained automaton is again minimal. By only changing the initial state of any minimal DFA, we keep a DFA that is complete and reduced. Furthermore, the obtained DFA is still accessible since we have seen in the previous paragraph that there is a path from the class of $(0, B)$ to the class of $(0, T)$, which is precisely the initial state in $\mathcal{M}_{m,r,\mathcal{T},2^p}$.

It remains to show that the language L accepted from the class of $(0, B)$ in the automaton $\mathcal{M}_{m,r,\mathcal{T},2^p}$ is equal to $\mathrm{val}_{2^p}^{-1}(m\overline{\mathcal{T}} + r)$. By construction, L is equal to the language

$L_{(0,B)}$ accepted from the state $(0, B)$ in the automaton $\Pi\left(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}\right)$ and we already know that $L_{(0,T)} = \mathrm{val}_{2^p}^{-1}(m\mathcal{T} + r)$.

Let $w \in A_{2^p}$. We know that $w \in L_{(0,B)} \iff \mathrm{rep}_{2^p}(m)w \in L_{(0,T)}$. Thus, it is sufficient to prove that $\mathrm{val}_{2^p}(w) \in m\overline{\mathcal{T}} + r \iff m2^{p|w|} + \mathrm{val}_{2^p}(w) \in m\mathcal{T} + r$. In both cases, we must have that $\mathrm{val}_{2^p}(w) = mq + r$ with $q \in \mathbb{N}$. Since $q \leqslant \mathrm{val}_{2^p}(w) < 2^{p|w|}$, we have $\mathrm{rep}_2(2^{p|w|} + q) = 10^{p|w| - |\mathrm{rep}_2(q)|}\mathrm{rep}_2(q)$. This shows that $q \in \overline{\mathcal{T}} \iff 2^{p|w|} + q \in \mathcal{T}$, hence the conclusion. $\qquad\square$

Corollary 51. *Let m, p be positive integers and $r \in [\![0, m-1]\!]$. Then the state complexity of $m\overline{\mathcal{T}} + r$ with respect to the base 2^p is equal to $2k + \left\lceil \frac{z}{p} \right\rceil$ if $m = k2^z$ with k odd.*

9 Conclusion and perspectives

Our method is constructive and in principle, it may be applied to any b-recognizable set $X \subseteq \mathbb{N}$. However, in general, it is not the case that the product automaton $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{X,2^p}$ recognizing the bidimensional set $\{(n, mn + r) : n \in X\}$ is minimal. As an example, consider the 2-recognizable set X of powers of 2: $X = \{2^n : n \in \mathbb{N}\}$. The product automaton $\mathcal{A}_{3,0,2} \times \mathcal{A}_{X,2}$ of our construction (for $m = 3$, $r = 0$ and $b = 2$) has 6 states but is clearly not minimal since it is easily checked that the automaton of Figure 15 is the trim minimal automaton recognizing the set $\{(2^n, 3 \cdot 2^n) : n \in \mathbb{N}\}$. This illustrates
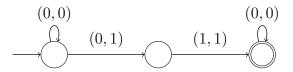

Figure 15: Minimal automaton recognizing the set $\{(2^n, 3 \cdot 2^n) : n \in \mathbb{N}\}$.

that, in general, the minimization procedure is not only needed in the final projection $\Pi\left(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{X,2^p}\right)$ as is the case in the present work.

Nevertheless, we conjecture that the phenomenon described in this work for the Thue-Morse set also appears for all b-recognizable sets of the form

$$X_{b,c,M,R} = \{n \in \mathbb{N} : |\mathrm{rep}_b(n)|_c \equiv R \bmod M\}$$

where b is an integer base, c is any digit in A_b, M is an integer greater than or equal to 2 and R is any possible remainder in $[\![0, M-1]\!]$. More precisely, we conjecture that whenever the base b is a prime power, i.e. $b = q^p$ for some prime q, then the state complexity of $mX_{b,c,M,R} + r$ is given by the formula $Mk + \lceil\frac{z}{p}\rceil$ where k is the part of the multiple m that is prime to the base b, i.e. $m = kq^z$ with $\gcd(k, q) = 1$. Note that the set $\overline{\mathcal{T}}$ is of this form: $\overline{\mathcal{T}} = \{n \in \mathbb{N} : |\mathrm{rep}_2(n)|_1 \equiv 0 \bmod 2\}$.

Another potential future research direction in the continuation of the present work is to consider automata reading the expansions of numbers with least significant digit first. Both reading directions are relevant to different problems. For example, it is easier to

compute addition thanks to an automaton reading expansions from "right to left" than from "left to right". On the opposite, if we have in mind to generalize our problems to b-recognizable sets of real numbers (see for instance [9, 13, 15]), then the relevant reading direction is the one with most significant digit first. Further, there is no intrinsic reason why the state complexity from "left to right" should be the same as (or even close to) that obtained from "right to left" since in general, it is well known that the state complexity of an arbitrary language can greatly differ from that of its reversed language, see for example [29].

References

[1] B. Alexeev. Minimal DFA for testing divisibility. *J. Comput. System Sci.*, 69(2):235–243, 2004.

[2] J.-P. Allouche. Thue, combinatorics on words, and conjectures inspired by the Thue-Morse sequence. *J. Théor. Nombres Bordeaux*, 27:375–388, 2015.

[3] J.-P. Allouche, B. Cloitre, and V. Shevelev. Beyond odious and evil. *Aequationes Math.*, 90(2):341–353, 2016.

[4] J.-P. Allouche, N. Rampersad, and J. Shallit. Periodicity, repetitions, and orbits of an automatic sequence. *Theoret. Comput. Sci.*, 410(30-32):2795–2803, 2009.

[5] J.-P. Allouche and J. Shallit. The ring of k-regular sequences. *Theoret. Comput. Sci.*, 98(2):163–197, 1992.

[6] J.-P. Allouche and J. Shallit. The ubiquitous Prouhet-Thue-Morse sequence. In *Sequences and their applications (Singapore, 1998)*, Springer Ser. Discrete Math. Theor. Comput. Sci., pages 1–16. Springer, London, 1999.

[7] J.-P. Allouche and J. Shallit. *Automatic sequences*. Cambridge University Press, Cambridge, 2003. Theory, applications, generalizations.

[8] B. Boigelot, I. Mainz, V. Marsault, and M. Rigo. An efficient algorithm to decide periodicity of b-recognisable sets using MSDF convention. In *44th International Colloquium on Automata, Languages, and Programming*, volume 80 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 118, 14. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2017.

[9] B. Boigelot, S. Rassart, and P. Wolper. On the expressiveness of real and integer arithmetic automata (extended abstract). In *ICALP*, volume 1443 of *Lecture Notes in Comput. Sci.*, pages 152–163. Springer, Berlin, 1998.

[10] V. Bruyère and G. Hansel. Bertrand numeration systems and recognizability. *Theoret. Comput. Sci.*, 181(1):17–43, 1997. Latin American Theoretical INformatics (Valparaíso, 1995).

[11] V. Bruyère, G. Hansel, C. Michaux, and R. Villemaire. Logic and p-recognizable sets of integers. *Bull. Belg. Math. Soc. Simon Stevin*, 1(2):191–238, 1994. Journées Montoises (Mons, 1992).

[12] M. Bucci, N. Hindman, S. Puzynina, and L. Q. Zamboni. On additive properties of sets defined by the Thue-Morse word. *J. Combin. Theory Ser. A*, 120(6):1235–1245, 2013.

[13] É. Charlier. First-order logic and numeration systems. In *Sequences, groups, and number theory*, Trends Math., pages 89–141. Birkhäuser/Springer, Cham, 2018.

[14] É. Charlier, C. Cisternino, and A. Massuir. State complexity of the multiples of the Thue-Morse set. In *Tenth International Symposium on Games, Automata, Logics, and Formal Verification (GandALF 2019)*, volume 305 of *Electronic Proceedings in Theoretical Computer Science*, pages 34–49. Open Publishing Association, 2019.

[15] É. Charlier, J. Leroy, and M. Rigo. An analogue of Cobham's theorem for graph directed iterated function systems. *Adv. Math.*, 280:86–120, 2015.

[16] É. Charlier and N. Rampersad. The growth function of S-recognizable sets. *Theoret. Comput. Sci.*, 412(39):5400–5408, 2011.

[17] É. Charlier, N. Rampersad, M. Rigo, and L. Waxweiler. The minimal automaton recognizing $m\mathbb{N}$ in a linear numeration system. *Integers*, 11B:Paper No. A4, 24, 2011.

[18] É. Charlier, N. Rampersad, and J. Shallit. Enumeration and decidable properties of automatic sequences. *Internat. J. Found. Comput. Sci.*, 23(5):1035–1066, 2012.

[19] A. Cobham. On the base-dependence of sets of numbers recognizable by finite automata. *Math. Systems Theory*, 3:186–192, 1969.

[20] A. Cobham. Uniform tag sequences. *Math. Systems Theory*, 6, 1972.

[21] F. Durand. Decidability of the HD0L ultimate periodicity problem. *RAIRO Theor. Inform. Appl.*, 47(2):201–214, 2013.

[22] S. Eilenberg. *Automata, languages, and machines. Vol. A*. Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York, 1974. Pure and Applied Mathematics, Vol. 58.

[23] C. Frougny. Representations of numbers and finite automata. *Math. Systems Theory*, 25(1):37–60, 1992.

[24] C. Frougny and J. Sakarovitch. Number representation and finite automata. In *Combinatorics, automata and number theory*, volume 135 of *Encyclopedia Math. Appl.*, pages 34–107. Cambridge Univ. Press, Cambridge, 2010.

[25] J. Honkala. A decision method for the recognizability of sets defined by number systems. *RAIRO Inform. Théor. Appl.*, 20(4):395–403, 1986.

[26] J. Leroux. A polynomial time Presburger criterion and synthesis for number decision diagrams. In *20th IEEE Symposium on Logic in Computer Science*, pages 147–156. IEEE Computer Society, Chicago, IL, USA, 2005.

[27] M. Lothaire. *Combinatorics on words*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1997.

[28] V. Marsault and J. Sakarovitch. Ultimate periodicity of b-recognisable sets: a quasilinear procedure. In *Developments in language theory*, volume 7907 of *Lecture Notes in Comput. Sci.*, pages 362–373. Springer, Heidelberg, 2013.

[29] V. Marsault. An efficient algorithm to decide periodicity of b-recognizable sets using LSDF convention. *Log. Methods Comput. Sci.*, 15(3), 2019.

[30] C. Mauduit. Multiplicative properties of the Thue-Morse sequence. *Period. Math. Hungar.*, 43(1-2):137–153, 2001.

[31] I. V. Mitrofanov. Periodicity of morphic words. *Fundam. Prikl. Mat.*, 18(4):107–119, 2013.

[32] H. Mousavi. Walnut. https://cs.uwaterloo.ca/~shallit/papers.html, 2015.

[33] A. A. Muchnik. The definable criterion for definability in Presburger arithmetic and its applications. *Theoret. Comput. Sci.*, 290(3):1433–1444, 2003.

[34] M. Queffélec. Questions around the Thue-Morse sequence. *Unif. Distrib. Theory*, 13(1):1–25, 2018.

[35] M. Rigo. *Formal languages, automata and numeration systems. 2.* Networks and Telecommunications Series. ISTE, London; John Wiley & Sons, Inc., Hoboken, NJ, 2014. Applications to recognizability and decidability, With a foreword by Valérie Berthé.

[36] J. Sakarovitch. *Elements of automata theory.* Cambridge University Press, Cambridge, 2009. Translated from the 2003 French original by Reuben Thomas.

[37] J. Shallit. Enumeration and automatic sequences. *Pure Math. Appl. (PU.M.A.)*, 25(1):96–106, 2015.

[38] L. Waxweiler. *Caractère reconnaissable d'ensembles de polynômes à coefficients dans un corps fini.* PhD thesis, University of Liège, Belgium, 2009.