Diverse Palindromic Factorization Is NP-complete

Hideo Bannai[1], Travis Gagie[2,3](✉), Shunsuke Inenaga[1], Juha Kärkkäinen[2], Dominik Kempa[2,3], Marcin Piątkowski[4], Simon J. Puglisi[2,3], and Shiho Sugimoto[1]

[1] Department of Informatics, Kyushu University, Fukuoka, Japan
[2] Department of Computer Science, University of Helsinki, Helsinki, Finland
[3] Helsinki Institute for Information Technology, Espoo, Finland
travis.gagie@gmail.com
[4] Faculty of Mathematics and Computer Science, Nicolaus Copernicus University, Toruń, Poland

Abstract. We prove that it is NP-complete to decide whether a given string can be factored into palindromes that are each unique in the factorization.

1 Introduction

Several papers have appeared on the subject of palindromic factorization. The palindromic length of a string is the minimum number of palindromic substrings into which the string can be factored. Notice that, since a single symbol is a palindrome, the palindromic length of a string is always defined and at most the length of the string. Ravsky [8] proved a tight bound on the maximum palindromic length of a binary string in terms of its length. Frid, Puzynina, and Zamboni [4] conjectured that any infinite string in which the palindromic length of any finite substring is bounded, is ultimately periodic. Their work led other researchers to consider how to efficiently compute a string's palindromic length and give a minimum palindromic factorization. It is not difficult to design a quadratic-time algorithm that uses linear space, but doing better than that seems to require some string combinatorics.

Alatabbi, Iliopoulos and Rahman [1] first gave a linear-time algorithm for computing a minimum factorization into maximal palindromes, if such a factorization exists. Notice that *abaca* cannot be factored into maximal palindromes, for example, because its maximal palindromes are *a*, *aba*, *a*, *aca* and *a*. Fici, Gagie, Kärkkäinen and Kempa [3] and I, Sugimoto, Inenaga, Bannai and Takeda [6] independently then described essentially the same $\mathcal{O}(n \log n)$-time

T. Gagie and S.J. Puglisi—Supported by grants 268324, 258308 and 284598 from the Academy of Finland.

M. Piątkowski—Supported by a research fellowship within the project "Enhancing Educational Potential of Nicolaus Copernicus University in the Disciplines of Mathematical and Natural Sciences" (project no. POKL.04.01.01-00-081/10).

I. Potapov (Ed.): DLT 2015, LNCS 9168, pp. 85–96, 2015.
DOI: 10.1007/978-3-319-21500-6_6

algorithm for computing a minimum palindromic factorization. Shortly thereafter, Kosolobov, Rubinchik and Shur [7] gave an algorithm for recognizing strings with a given palindromic length. Their result can be used to compute the palindromic length ℓ of a string of length n in $\mathcal{O}(n\ell \log \ell)$ time. We also note that Gawrychowski and Uznański [5] used similar techniques as Fici et al. and I et al., for finding approximately the longest palindrome in a stream.

We call a factorization *diverse* if each of the factors is unique. Some well-known factorizations, such as the LZ77 [10] and LZ78 [11] parses, are diverse (except that the last factor may have appeared before). Fernau, Manea, Mercaş and Schmid [2] very recently proved that it is NP-complete to determine whether a given string has a diverse factorization of size at least k. It seems natural to consider the problem of determining whether a given string has a diverse factorization into palindromes. For example, *bgikkpps* and *bgikpspk* each have exactly one such factorization — i.e., (b, g, i, kk, pp, s) and (b, g, i, $kpspk$), respectively — but *bgkpispk* has none. This problem is obviously in NP and in this paper we prove that it is NP-hard and, thus, NP-complete. Some people might dismiss as doubly useless a lower bound for a problem with no apparent application; nevertheless, we feel the proof is pretty (albeit somewhat intricate) and we would like to share it. We conjecture that it is also NP-complete to determine whether a given string has a palindromic factorization in which each factor appears at most a given number $k > 1$ times.

2 Outline

The circuit satisfiability problem was one of the first to be proven NP-complete and is often the first taught in undergraduate courses. It asks whether a given Boolean circuit C is satisfiable, i.e., has an assignment to its inputs that makes its single output true. We will show how to build, in time linear in the size of C, a string that has a diverse palindromic factorization if and only if C is satisfiable. It follows that diverse palindromic factorization is also NP-hard. Our construction is similar to the Tseitin Transform [9] from Boolean circuits to CNF formulas.

Because AND, OR and NOT gates can be implemented with a constant number of NAND gates, we assume without loss of generality that C is composed only of NAND gates with two inputs and one output each, and splitters that each divide one wire into two. Furthermore, we assume each wire in C is labelled with a unique symbol (considering a split to be the end of an incoming wire and the beginning of two new wires, so all three wires have different labels). For each such symbol a, and some auxiliary symbols we introduce during our construction, we use as characters in our construction three related symbols: a itself, $\bar{a}$ and x_a. We indicate an auxiliary symbol related to a by writing a' or a''. We write x_a^j to denote j copies of x_a. We emphasize that, despite their visual similarity, a and $\bar{a}$ are separate characters, which play complementary roles in our reduction. We use \$ and # as generic separator symbols, which we consider to be distinct for each use; to prevent confusion, we add different superscripts to their different uses within the same part of the construction.

We can build a sequence $C_0, \ldots, C_t$ of subcircuits such that C_0 is empty, $C_t = C$ and, for $1 \leq i \leq t$, we obtain C_i from C_{i-1} by one of the following operations:

– adding a new wire (which is both an input and an output in C_i),
– splitting an output of C_{i-1} into two outputs,
– making two outputs of C_{i-1} the inputs of a new NAND gate.

We will show how to build in time linear in the size of C, inductively and in turn, a sequence of strings $S_1, \ldots, S_t$ such that S_i represents C_i according to the following definitions:

Definition 1. *A diverse palindromic factorization P of a string S_i* encodes *an assignment τ to the inputs of a circuit C_i if the following conditions hold:*

– *if τ makes an output of C_i labelled a true, then a, x_a and $x_a \bar{a} x_a$ are complete factors in P but $\bar{a}$, $x_a a x_a$ and x_a^j are not for $j > 1$;*
– *if τ makes an output of C_i labelled a false, then $\bar{a}$, x_a and $x_a a x_a$ are complete factors in P but a, $x_a \bar{a} x_a$ and x_a^j are not for $j > 1$;*
– *if a is a label in C but not in C_i, then none of a, $\bar{a}$, $x_a a x_a$, $x_a \bar{a} x_a$ and x_a^j for $j \geq 1$ are complete factors in P.*

Definition 2. *A string S_i* represents *a circuit C_i if each assignment to the inputs of C_i is encoded by some diverse palindromic factorization of S_i, and each diverse palindromic factorization of S_i encodes some assignment to the inputs of C_i.*

Once we have S_t, we can easily build in constant time a string S that has a diverse palindromic factorization if and only if C is satisfiable. To do this, we append $\$\#\, x_a a x_a$ to S_t, where \$ and # are symbols not occurring in S_t and a is the label on C's output. Since \$ and # do not occur in S_t and occur as a pair of consecutive characters in S, they must each be complete factors in any palindromic factorization of S. It follows that there is a diverse palindromic factorization of S if and only if there is a diverse palindromic factorization of S_t in which $x_a a x_a$ is not a factor, which is the case if and only if there is an assignment to the inputs of C that makes its output true.

3 Adding a Wire

Suppose C_i is obtained from C_{i-1} by adding a new wire labelled a. If $i = 1$ then we set $S_i = x_a a x_a \bar{a} x_a$, whose two diverse palindromic factorizations $(x_a,\ a,\ x_a \bar{a} x_a)$ and $(x_a a x_a,\ \bar{a},\ x_a)$ encode the assignments true and false to the wire labelled a, which is both the input and output in C_i. If $i > 1$ then we set

$$S_i = S_{i-1}\,\$\#\, x_a a x_a \bar{a} x_a\,,$$

where \$ and # are symbols not occurring in S_{i-1} and not equal to a', $\overline{a'}$ or $x_{a'}$ for any label a' in C.

Since \$ and # do not occur in S_{i-1} and occur as a pair of consecutive characters in S_i, they must each be complete factors in any palindromic factorization of S_i. Therefore, any diverse palindromic factorization of S_i is the concatenation of a diverse palindromic factorization of S_{i-1} and either (\$, #, x_a, a, $x_a\bar{a}x_a$) or (\$, #, $x_a a x_a$, $\bar{a}$, x_a). Conversely, any diverse palindromic factorization of S_{i-1} can be extended to a diverse palindromic factorization of S_i by appending either (\$, #, x_a, a, $x_a\bar{a}x_a$) or (\$, #, $x_a a x_a$, $\bar{a}$, x_a).

Assume S_{i-1} represents C_{i-1}. Let τ be an assignment to the inputs of C_i and let P be a diverse palindromic factorization of S_{i-1} encoding τ restricted to the inputs of C_{i-1}. If τ makes the input (and output) of C_i labelled a true, then P concatenated with (\$, #, x_a, a, $x_a\bar{a}x_a$) is a diverse palindromic factorization of S_i that encodes τ. If τ makes that input false, then P concatenated with (\$, #, $x_a a x_a$, $\bar{a}$, x_a) is a diverse palindromic factorization of S_i that encodes τ. Therefore, each assignment to the inputs of C_i is encoded by some diverse palindromic factorization of S_i.

Now let P be a diverse palindromic factorization of S_i and let τ be the assignment to the inputs of C_{i-1} that is encoded by a prefix of P. If P ends with (\$, #, x_a, a, $x_a\bar{a}x_a$) then P encodes the assignment to the inputs of C_i that makes the input labelled a true and makes the other inputs true or false according to τ. If P ends with (\$, #, $x_a a x_a$, $\bar{a}$, x_a) then P encodes the assignment to the inputs of C_i that makes the input labelled a false and makes the other inputs true or false according to τ. Therefore, each diverse palindromic factorization of S_i encodes some assignment to the inputs of C_i.

Lemma 1. *We can build a string S_1 that represents C_1. If we have a string S_{i-1} that represents C_{i-1} and C_i is obtained from C_{i-1} by adding a new wire, then in constant time we can append symbols to S_{i-1} to obtain a string S_i that represents C_i.*

4 Splitting a Wire

Now suppose C_i is obtained from C_{i-1} by splitting an output of C_{i-1} labelled a into two outputs labelled b and c. We set

$$S'_i = S_{i-1}\,\$\#\,x_a^3 b' x_a a x_a c' x_a^5\,\$'\#'\,x_a^7 \overline{b'} x_a \bar{a} x_a \overline{c'} x_a^9\,,$$

where \$, \$', #, #', b', $\overline{b'}$, c' and $\overline{c'}$ are symbols not occurring in S_{i-1} and not equal to a', $\overline{a'}$ or $x_{a'}$ for any label a' in C.

Since \$, \$', # and #' do not occur in S_{i-1} and occur as pairs of consecutive characters in S'_i, they must each be complete factors in any palindromic factorization of S'_i. Therefore, a simple case analysis shows that any diverse palindromic factorization of S'_i is the concatenation of a diverse palindromic factorization of S_{i-1} and one of

$$(\$,\ \#,\ x_a^3,\ b',\ x_a a x_a,\ c',\ x_a^5,\ \$',\ \#',\ x_a^2,\ x_a^4,\ x_a \overline{b'} x_a,\ \bar{a},\ x_a \overline{c'} x_a,\ x_a^8)\,,$$
$$(\$,\ \#,\ x_a^3,\ b',\ x_a a x_a,\ c',\ x_a^5,\ \$',\ \#',\ x_a^4,\ x_a^2,\ x_a \overline{b'} x_a,\ \bar{a},\ x_a \overline{c'} x_a,\ x_a^8)\,,$$
$$(\$,\ \#,\ x_a^3,\ b',\ x_a a x_a,\ c',\ x_a^5,\ \$',\ \#',\ x_a^6,\ x_a \overline{b'} x_a,\ \bar{a},\ x_a \overline{c'} x_a,\ x_a^8)\,,$$
$$(\$,\ \#,\ x_a^2,\ x_a b' x_a,\ a,\ x_a c' x_a,\ x_a^4,\ \$',\ \#',\ x_a^7,\ \overline{b'},\ x_a \bar{a} x_a,\ \overline{c'},\ x_a^3,\ x_a^6)\,,$$
$$(\$,\ \#,\ x_a^2,\ x_a b' x_a,\ a,\ x_a c' x_a,\ x_a^4,\ \$',\ \#',\ x_a^7,\ \overline{b'},\ x_a \bar{a} x_a,\ \overline{c'},\ x_a^6,\ x_a^3)\,,$$
$$(\$,\ \#,\ x_a^2,\ x_a b' x_a,\ a,\ x_a c' x_a,\ x_a^4,\ \$',\ \#',\ x_a^7,\ \overline{b'},\ x_a \bar{a} x_a,\ \overline{c'},\ x_a^9)\,.$$

In any diverse palindromic factorization of S_i', therefore, either b' and c' are complete factors but $\overline{b'}$ and $\overline{c'}$ are not, or vice versa.

Conversely, any diverse palindromic factorization of S_{i-1} in which a, x_a and $x_a \bar{a} x_a$ are complete factors but $\bar{a}$, $x_a a x_a$ and x_a^j are not for $j > 1$, can be extended to a diverse palindromic factorization of S_i' by appending either of

$$(\$,\ \#,\ x_a^3,\ b',\ x_a a x_a,\ c',\ x_a^5,\ \$',\ \#',\ x_a^2,\ x_a^4,\ x_a \overline{b'} x_a,\ \bar{a},\ x_a \overline{c'} x_a,\ x_a^8)\,,$$
$$(\$,\ \#,\ x_a^3,\ b',\ x_a a x_a,\ c',\ x_a^5,\ \$',\ \#',\ x_a^6,\ x_a \overline{b'} x_a,\ \bar{a},\ x_a \overline{c'} x_a,\ x_a^8)\,;$$

any diverse palindromic factorization of S_{i-1} in which $\bar{a}$, x_a and $x_a a x_a$ are complete factors but a, $x_a \bar{a} x_a$ and x_a^j are not for $j > 1$, can be extended to a diverse palindromic factorization of S_i' by appending either of

$$(\$,\ \#,\ x_a^2,\ x_a b' x_a,\ a,\ x_a c' x_a,\ x_a^4,\ \$',\ \#',\ x_a^7,\ \overline{b'},\ x_a \bar{a} x_a,\ \overline{c'},\ x_a^3,\ x_a^6)\,,$$
$$(\$,\ \#,\ x_a^2,\ x_a b' x_a,\ a,\ x_a c' x_a,\ x_a^4,\ \$',\ \#',\ x_a^7,\ \overline{b'},\ x_a \bar{a} x_a,\ \overline{c'},\ x_a^9)\,.$$

We set

$$S_i = S_i'\,\$''\#''\,x_b b x_b b' x_b \overline{b'} x_b \bar{b} x_b\,\$'''\#'''\,x_c c x_c c' x_c \overline{c'} x_c \bar{c} x_c\,,$$

where $\$''$, $\$'''$, $\#''$ and $\#'''$ are symbols not occurring in S_i' and not equal to a', $\overline{a'}$ or $x_{a'}$ for any label a' in C. Since $\$''$, $\$'''$, $\#''$ and $\#'''$ do not occur in S_i' and occur as pairs of consecutive characters in S_i', they must each be complete factors in any palindromic factorization of S_i. Therefore, any diverse palindromic factorization of S_i is the concatenation of a diverse palindromic factorization of S_i' and one of

$$(\$'',\ \#'',\ x_b,\ b,\ x_b b' x_b,\ \overline{b'},\ x_b \bar{b} x_b,\ \$''',\ \#''',\ x_c,\ c,\ x_c c' x_c,\ \overline{c'},\ x_c \bar{c} x_c)\,,$$
$$(\$'',\ \#'',\ x_b b x_b,\ b',\ x_b \overline{b'} x_b,\ \bar{b},\ x_b,\ \$''',\ \#''',\ x_c c x_c,\ c',\ x_c \overline{c'} x_c, \bar{c}, x_c)\,.$$

Conversely, any diverse palindromic factorization of S_i' in which b' and c' are complete factors but $\overline{b'}$ and $\overline{c'}$ are not, can be extended to a diverse palindromic factorization of S_i by appending

$$(\$'',\ \#'',\ x_b,\ b,\ x_b b' x_b,\ \overline{b'},\ x_b \bar{b} x_b,\ \$''',\ \#''',\ x_c,\ c,\ x_c c' x_c,\ \overline{c'},\ x_c \bar{c} x_c)\,;$$

any diverse palindromic factorization of S_i' in which $\overline{b'}$ and $\overline{c'}$ are complete factors but b' and c' are not, can be extended to a diverse palindromic factorization of S_i by appending

$$(\$'',\ \#'',\ x_b b x_b,\ b',\ x_b \overline{b'} x_b,\ \bar{b},\ x_b,\ \$''',\ \#''',\ x_c c x_c,\ c',\ x_c \overline{c'} x_c,\ \bar{c},\ x_c)\,.$$

Assume S_{i-1} represents C_{i-1}. Let τ be an assignment to the inputs of C_{i-1} and let P be a diverse palindromic factorization of S_{i-1} encoding τ. If τ makes the output of C_{i-1} labelled a true, then P concatenated with, e.g.,

$$(\$,\ \#,\ x_a^3,\ b',\ x_a a x_a,\ c',\ x_a^5,\ \$',\ \#',\ x_a^2,\ x_a^4,\ x_a\overline{b'}x_a,\ \bar{a},\ x_a\overline{c'}x_a,\ x_a^8,\ \$'',\ \#'',\ x_b,\ b,\ x_b b' x_b,\ \overline{b'},\ x_b\bar{b}x_b,\ \$''',\ \#''',\ x_c,\ c,\ x_c c' x_c,\ \overline{c'},\ x_c\bar{c}x_c)$$

is a diverse palindromic factorization of S_i. Notice b, c, x_b, x_c, $x_b\bar{b}x_b$ and $x_c\bar{c}x_c$ are complete factors but $\bar{b}$, $\bar{c}$, $x_b b x_b$, $x_c c x_c$, x_b^j and x_c^j for $j > 1$ are not. Therefore, this concatenation encodes the assignment to the inputs of C_i that makes them true or false according to τ.

If τ makes the output of C_{i-1} labelled a false, then P concatenated with, e.g.,

$$(\$,\ \#,\ x_a^2,\ x_a b' x_a,\ a,\ x_a c' x_a,\ x_a^4,\ \$',\ \#',\ x_a^7,\ \overline{b'},\ x_a\bar{a}x_a,\ \overline{c'},\ x_a^3,\ x_a^6,\ \$'',\ \#'',\ x_b b x_b,\ b',\ x_b\overline{b'}x_b,\ \bar{b},\ x_b,\ \$''',\ \#''',\ x_c c x_c,\ c',\ x_c\overline{c'}x_c,\ \bar{c},\ x_c)$$

is a diverse palindromic factorization of S_i. Notice $\bar{b}$, $\bar{c}$, x_b, x_c, $x_b b x_b$ and $x_c c x_c$ are complete factors but b, c, $x_b\bar{b}x_b$, $x_c\bar{c}x_c$, x_b^j and x_c^j for $j > 1$ are not. Therefore, this concatenation encodes the assignment to the inputs of C_i that makes them true or false according to τ. Since C_{i-1} and C_i have the same inputs, each assignment to the inputs of C_i is encoded by some diverse palindromic factorization of S_i.

Now let P be a diverse palindromic factorization of S_i and let τ be the assignment to the inputs of C_{i-1} that is encoded by a prefix of P. If P ends with any of

$$(\$,\ \#,\ x_a^3,\ b',\ x_a a x_a,\ c',\ x_a^5,\ \$',\ \#',\ x_a^2,\ x_a^4,\ x_a\overline{b'}x_a,\ \bar{a},\ x_a\overline{c'}x_a,\ x_a^8),$$
$$(\$,\ \#,\ x_a^3,\ b',\ x_a a x_a,\ c',\ x_a^5,\ \$',\ \#',\ x_a^4,\ x_a^2,\ x_a\overline{b'}x_a,\ \bar{a},\ x_a\overline{c'}x_a,\ x_a^8),$$
$$(\$,\ \#,\ x_a^3,\ b',\ x_a a x_a,\ c',\ x_a^5,\ \$',\ \#',\ x_a^6,\ x_a\overline{b'}x_a,\ \bar{a},\ x_a\overline{c'}x_a,\ x_a^8)$$

followed by

$$(\$'',\ \#'',\ x_b,\ b,\ x_b b' x_b,\ \overline{b'},\ x_b\bar{b}x_b,\ \$''',\ \#''',\ x_c,\ c,\ x_c c' x_c,\ \overline{c'},\ x_c\bar{c}x_c),$$

then a must be a complete factor in the prefix of P encoding τ, so τ must make the output of C_{i-1} labelled a true. Since b, c, x_b, x_c, $x_b\bar{b}x_b$ and $x_c\bar{c}x_c$ are complete factors in P but $\bar{b}$, $\bar{c}$, $x_b b x_b$, $x_c c x_c$, x_b^j and x_c^j for $j > 1$ are not, P encodes the assignment to the inputs of C_i that makes them true or false according to τ.

If P ends with any of

$$(\$,\ \#,\ x_a^2,\ x_a b' x_a,\ a,\ x_a c' x_a,\ x_a^4,\ \$',\ \#',\ x_a^7,\ \overline{b'},\ x_a\bar{a}x_a,\ \overline{c'},\ x_a^3,\ x_a^6),$$
$$(\$,\ \#,\ x_a^2,\ x_a b' x_a,\ a,\ x_a c' x_a,\ x_a^4,\ \$',\ \#',\ x_a^7,\ \overline{b'},\ x_a\bar{a}x_a,\ \overline{c'},\ x_a^6,\ x_a^3),$$
$$(\$,\ \#,\ x_a^2,\ x_a b' x_a,\ a,\ x_a c' x_a,\ x_a^4,\ \$',\ \#',\ x_a^7,\ \overline{b'},\ x_a\bar{a}x_a,\ \overline{c'},\ x_a^9)$$

followed by

$$(\$'',\ \#'',\ x_b b x_b,\ b',\ x_b\overline{b'}x_b,\ \bar{b},\ x_b,\ \$''',\ \#''',\ x_c c x_c,\ c',\ x_c\overline{c'}x_c, \bar{c}, x_c),$$

then $\bar{a}$ must be a complete factor in the prefix of P encoding τ, so τ must make the output of C_{i-1} labelled a false. Since $\bar{b}$, $\bar{c}$, x_b, x_c, $x_b b x_b$ and $x_c c x_c$ are complete factors but b, c, $x_b \bar{b} x_b$, $x_c \bar{c} x_c$, x_b^j and x_c^j for $j > 1$ are not, P encodes the assignment to the inputs of C_i that makes them true or false according to τ.

Since these are all the possibilities for how P can end, each diverse palindromic factorization of S_i encodes some assignment to the inputs of C_i. This gives us the following lemma:

Lemma 2. *If we have a string S_{i-1} that represents C_{i-1} and C_i is obtained from C_{i-1} by splitting an output of C_{i-1} into two outputs, then in constant time we can append symbols to S_{i-1} to obtain a string S_i that represents C_i.*

5 Adding a NAND Gate

Finally, suppose C_i is obtained from C_{i-1} by making two outputs of C_{i-1} labelled a and b the inputs of a new NAND gate whose output is labelled c. Let C'_{i-1} be the circuit obtained from C_{i-1} by splitting the output of C_{i-1} labelled a into two outputs labelled a_1 and a_2, where a_1 and a_2 are symbols we use only here. Assuming S_{i-1} represents C_{i-1}, we can use Lemma 2 to build in constant time a string S'_{i-1} representing C'_{i-1}. We set

$$\begin{aligned} S'_i = S'_{i-1}\,\$\#\, & x_{c'}^3 a'_1 x_{c'} a_1 x_{c'} \overline{a_1} x_{c'} \overline{a'_1} x_{c'}^5 \\ & \$'\#'\, x_{c'}^7 a'_2 x_{c'} a_2 x_{c'} \overline{a_2} x_{c'} \overline{a'_2} x_{c'}^9 \\ & \$''\#''\, x_{c'}^{11} b' x_{c'} b x_{c'} \bar{b} x_{c'} \overline{b'} x_{c'}^{13}\,, \end{aligned}$$

where all of the symbols in the suffix after S'_{i-1} are ones we use only here.

Since \$, \$′, \$″, \$‴, # and #′ do not occur in S_{i-1} and occur as pairs of consecutive characters in S'_i, they must each be complete factors in any palindromic factorization of S'_i. Therefore, any diverse palindromic factorization of S'_i consists of

1. a diverse palindromic factorization of S'_{i-1},
2. (\$, #),
3. a diverse palindromic factorization of $x_{c'}^3 a'_1 x_{c'} a_1 x_{c'} \overline{a_1} x_{c'} \overline{a'_1} x_{c'}^5$,
4. (\$′, #′),
5. a diverse palindromic factorization of $x_{c'}^7 a'_2 x_{c'} a_2 x_{c'} \overline{a_2} x_{c'} \overline{a'_2} x_{c'}^9$,
6. (\$″, #″),
7. a diverse palindromic factorization of $x_{c'}^{11} b' x_{c'} b x_{c'} \bar{b} x_{c'} \overline{b'} x_{c'}^{13}$.

If a_1 is a complete factor in the factorization of S'_{i-1}, then the diverse palindromic factorization of

$$x_{c'}^3 a'_1 x_{c'} a_1 x_{c'} \overline{a_1} x_{c'} \overline{a'_1} x_{c'}^5$$

must include either

$$(a'_1,\ x_{c'} a_1 x_{c'},\ \overline{a_1},\ x_{c'} \overline{a'_1} x_{c'}) \qquad \text{or} \qquad (a'_1,\ x_{c'} a_1 x_{c'},\ \overline{a_1},\ x_{c'},\ \overline{a'_1})\,.$$

Notice that in the former case, the factorization need not contain $x_{c'}$. If $\overline{a_1}$ is a complete factor in the factorization of S'_{i-1}, then the diverse palindromic factorization of

$$x_{c'}^3 a'_1 x_{c'} a_1 x_{c'} \overline{a_1} x_{c'} \overline{a'_1} x_{c'}^5$$

must include either

$$(x_{c'} a'_1 x_{c'},\ a_1,\ x_{c'} \overline{a_1} x_{c'},\ \overline{a'_1}) \qquad \text{or} \qquad (a'_1,\ x_{c'},\ a_1,\ x_{c'} \overline{a_1} x_{c'},\ \overline{a'_1})\,.$$

Again, in the former case, the factorization need not contain $x_{c'}$. A simple case analysis shows analogous propositions hold for a_2 and b; we leave the details for the full version of this paper.

We set

$$S''_i = S'_i\ \$^\dagger \#^\dagger\ x_{c'}^{15} \overline{a'_1} x_{c'} c' x_{c'} \overline{b'} x_{c'}^{17}\ \$^{\dagger\dagger} \#^{\dagger\dagger}\ x_{c'}^{19} \overline{a'_2} x_{c'} d x_{c'} b' x_{c'}^{21}\,,$$

where $\$^\dagger$, $\#^\dagger$, $\$^{\dagger\dagger}$, $\#^{\dagger\dagger}$, c' and d are symbols we use only here. Any diverse palindromic factorization of S''_i consists of

1. a diverse palindromic factorization of S'_i,
2. $(\$^\dagger,\ \#^\dagger)$,
3. a diverse palindromic factorization of $x_{c'}^{15} \overline{a'_1} x_{c'} c' x_{c'} \overline{b'} x_{c'}^{17}$,
4. $(\$^{\dagger\dagger},\ \#^{\dagger\dagger})$,
5. a diverse palindromic factorization of $x_{c'}^{19} \overline{a'_2} x_{c'} d x_{c'} b' x_{c'}^{21}$.

Since a_1 and a_2 label outputs in C'_{i-1} split from the same output in C_{i-1}, it follows that a_1 is a complete factor in a diverse palindromic factorization of S'_{i-1} if and only if a_2 is. Therefore, we need consider only four cases:

- The factorization of S'_{i-1} includes a_1, a_2 and b as complete factors, so the factorization of S'_i includes as complete factors either $x_{c'} \overline{a'_1} x_{c'}$, or $\overline{a'_1}$ and $x_{c'}$; either $x_{c'} \overline{a'_2} x_{c'}$, or $\overline{a'_2}$ and $x_{c'}$; either $x_{c'} \overline{b'} x_{c'}$, or $\overline{b'}$ and $x_{c'}$; and b'. Trying all the combinations — there are only four, since $x_{c'}$ can appear as a complete factor at most once — shows that any diverse palindromic factorization of S''_i includes one of

$$(\overline{a'_1},\ x_{c'} c' x_{c'},\ \overline{b'},\ \ldots,\ \overline{a'_2},\ x_{c'},\ d,\ x_{c'} b' x_{c'})\,,$$
$$(\overline{a'_1},\ x_{c'} c' x_{c'},\ \overline{b'},\ \ldots,\ x_{c'} \overline{a'_2} x_{c'},\ d,\ x_{c'} b' x_{c'})\,,$$

  with the latter only possible if $x_{c'}$ appears earlier in the factorization.
- The factorization of S'_{i-1} includes a_1, a_2 and $\overline{b}$ as complete factors, so the factorization of S'_i includes as complete factors either $x_{c'} \overline{a'_1} x_{c'}$, or $\overline{a'_1}$ and $x_{c'}$; either $x_{c'} \overline{a'_2} x_{c'}$, or $\overline{a'_2}$ and $x_{c'}$; $\overline{b'}$; and either $x_{c'} b' x_{c'}$, or b' and $x_{c'}$. Trying all the combinations shows that any diverse palindromic factorization of S''_i includes one of

$$(\overline{a'_1},\ x_{c'},\ c',\ x_{c'} \overline{b'} x_{c'},\ \ldots,\ \overline{a'_2},\ x_{c'} d x_{c'},\ b')\,,$$
$$(x_{c'} \overline{a'_1} x_{c'},\ c',\ x_{c'} \overline{b'} x_{c'},\ \ldots,\ \overline{a'_2},\ x_{c'} d x_{c'},\ b')\,,$$

  with the latter only possible if $x_{c'}$ appears earlier in the factorization.

– The factorization of S'_{i-1} includes $\overline{a_1}$, $\overline{a_2}$ and b as complete factors, so the factorization of S'_i includes as complete factors $\overline{a'_1}$; $\overline{a'_2}$; either $x_{c'}\overline{b'}x_{c'}$, or $\overline{b'}$ and $x_{c'}$; and b'. Trying all the combinations shows that any diverse palindromic factorization of S''_i includes one of

$$(x_{c'}\overline{a'_1}x_{c'},\ c',\ x_{c'},\ \overline{b'},\ \ldots,\ x_{c'}\overline{a'_2}x_{c'},\ d,\ x_{c'}b'x_{c'})\,,$$
$$(x_{c'}\overline{a'_1}x_{c'},\ c',\ x_{c'}\overline{b'}x_{c'},\ \ldots,\ x_{c'}\overline{a'_2}x_{c'},\ d,\ x_{c'}b'x_{c'})\,,$$

with the latter only possible if $x_{c'}$ appears earlier in the factorization.

– The factorization of S'_{i-1} includes $\overline{a_1}$, $\overline{a_2}$ and $\overline{b}$ as complete factors, so the factorization of S'_i includes as complete factors $\overline{a'_1}$; $\overline{a'_2}$; $\overline{b'}$; and either $x_{c'}b'x_{c'}$, or b' and $x_{c'}$. Trying all the combinations shows that any diverse palindromic factorization of S''_i that extends the factorization of S'_i includes one of

$$(x_{c'}\overline{a'_1}x_{c'},\ c',\ x_{c'}\overline{b'}x_{c'},\ \ldots,\ x_{c'}\overline{a'_2}x_{c'},\ d,\ x_{c'},\ b)\,,$$
$$(x_{c'}\overline{a'_1}x_{c'},\ c',\ x_{c'}\overline{b'}x_{c'},\ \ldots,\ x_{c'}\overline{a'_2}x_{c'},\ d,\ x_{c'}b'x_{c'})\,,$$

with the latter only possible if $x_{c'}$ appears earlier in the factorization.

Summing up, any diverse palindromic factorization of S''_i always includes $x_{c'}$ and includes either $x_{c'}c'x_{c'}$ if the factorization of S'_{i-1} includes a_1, a_2 and b as complete factors, or c' otherwise.

We set

$$S'''_i = S''_i\,\$^{\dagger\dagger\dagger}\#^{\dagger\dagger\dagger}\,x_{c'}^{23}c''x_{c'}c'x_{c'}\overline{c'}x_{c'}\overline{c''}x_{c'}^{25}\,,$$

where $\$^{\dagger\dagger\dagger}$ and $\#^{\dagger\dagger\dagger}$ are symbols we use only here. Any diverse palindromic factorization of S'''_i consists of

1. a diverse palindromic factorization of S''_i,
2. $(\$^{\dagger\dagger\dagger},\ \#^{\dagger\dagger\dagger})$,
3. a diverse palindromic factorization of $x_{c'}^{23}c''x_{c'}c'x_{c'}\overline{c'}x_{c'}\overline{c''}x_{c'}^{25}$.

Since $x_{c'}$ must appear as a complete factor in the factorization of S''_i, if c' is a complete factor in the factorization of S''_i, then the factorization of

$$x_{c'}^{23}\overline{c''}x_{c'}c'x_{c'}\overline{c'}x_{c'}c''x_{c'}^{25}$$

must include

$$(c'',\ x_{c'}c'x_{c'},\ \overline{c'},\ x_{c'}\overline{c''}x_{c'})\,;$$

otherwise, it must include

$$(x_{c'}c''x_{c'},\ c',\ x_{c'}\overline{c'}x_{c'},\ \overline{c''})\,.$$

That is, the factorization of $x_{c'}^{23}\overline{c''}x_{c'}c'x_{c'}\overline{c'}x_{c'}c''x_{c'}^{25}$ includes c'', $x_{c'}$ and $x_{c'}\overline{c''}x_{c'}$ but not $\overline{c''}$ or $x_{c'}c''x_{c'}$, if and only if the factorization of S''_i includes c'; otherwise, it includes $\overline{c''}$, $x_{c'}$ and $x_{c'}c''x_{c'}$ but not c'' or $x_{c'}\overline{c''}x_{c'}$.

We can slightly modify and apply the results in Section 4 to build in constant time a string T such that in any diverse palindromic factorization of

$$S_i = S'''_i\,\$^{\ddagger}\#^{\ddagger}\,T\,,$$

if c'' is a complete factor in the factorization of S''', then c, x_c and $x_c\bar{c}x_c$ are complete factors in the factorization of T but $\bar{c}$, $x_c c x_c$ and x_c^j are not for $j > 1$; otherwise, $\bar{c}$, x_c and $x_c c x_c$ are complete factors but c, $x_c\bar{c}x_c$ and x_c^j are not for $j > 1$. Again, we leave the details for the full version of this paper.

Assume S_{i-1} represents C_{i-1}. Let τ be an assignment to the inputs of C_{i-1} and let P be a diverse palindromic factorization of S_{i-1} encoding τ. By Lemma 2 we can extend P to P' so that it encodes the assignment to the inputs of C'_{i-1} that makes them true or false according to τ. Suppose τ makes the output of C_{i-1} labelled a true but the output labelled b false. Then P' concatenated with, e.g.,

$$\begin{array}{c}(\$,\ \#,\ x_{c'}^3,\ a_1',\ x_{c'}a_1x_{c'},\ \overline{a_1},\ x_{c'}\overline{a_1'}x_{c'},\ x_c^4,\\ \$',\ \#',\ x_{c'}^7,\ a_2',\ x_{c'}a_2x_{c'},\overline{a_2},\ x_{c'}\overline{a_2'}x_{c'},\ x_{c'}^8,\\ \$'',\ \#'',\ x_{c'}^{10},\ x_{c'}b'x_{c'},\ b,\ x_{c'}\bar{b}x_{c'},\ \overline{b'},\ x_{c'}^{13})\end{array}$$

is a diverse palindromic factorization P'' of S_i' which, concatenated with, e.g.,

$$\begin{array}{c}(\$^\dagger,\ \#^\dagger,\ x_{c'}^{15},\ \overline{a_1'},\ x_{c'},\ c',\ x_{c'}\overline{b'}x_{c'},\ x_{c'}^{16},\\ \$^\ddagger,\ \#^\ddagger,\ x_{c'}^{19},\ \overline{a_2'},\ x_{c'}dx_{c'},\ b',\ x_{c'}^{21})\end{array}$$

is a diverse palindromic factorization P''' of S_i'' which, concatenated with, e.g.,

$$(\$^{\dagger\dagger\dagger},\ \#^{\dagger\dagger\dagger},\ x_{c'}^{23},\ \overline{c''},\ x_{c'}c'x_{c'},\ \overline{c'},\ x_{c'}c''x_{c'},\ x_{c'}^{24})$$

is a diverse palindromic factorization $P^\dagger$ of S_i'''. Since $P^\dagger$ does not contain c'' as a complete factor, it can be extended to a diverse palindromic factorization $P^\ddagger$ of S_i in which $\bar{c}$, x_c and $x_c c x_c$ are complete factors but c, $x_c\bar{c}x_c$ and x_c^j are not for $j > 1$. Notice $P^\ddagger$ encodes the assignment to the inputs of C_i that makes them true or false according to τ. The other three cases — in which τ makes the outputs labelled a and b both false, false and true, and both true — are similar and we leave them for the full version of this paper. Since C_{i-1} and C_i have the same inputs, each assignment to the inputs of C_i is encoded by some diverse palindromic factorization of S_i.

Now let P be a diverse palindromic factorization of S_i and let τ be the assignment to the inputs of C_{i-1} that is encoded by a prefix of P. Let P' be the prefix of P that is a diverse palindromic factorization of S_i''' and suppose the factorization of

$$x_{c'}^{23}c''x_{c'}c'x_{c'}\overline{c'}x_{c'}\overline{c''}x_{c'}^{25}$$

in P' includes $\overline{c''}$ as a complete factor, which is the case if and only if P includes $\bar{c}$, x_c and $x_c c x_c$ as complete factors but not c, $x_c\bar{c}x_c$ and x_c^j for $j > 1$. We will show that τ must make the outputs of C_{i-1} labelled a and b true. The other case — in which the factorization includes c'' as a complete factor and we want to show τ makes at least one of the inputs labelled a and b false — is similar but longer, and we leave it for the full version of this paper.

Let P'' be the prefix of P' that is a diverse palindromic factorization of S_i''. Since $\overline{c''}$ is a complete factor in the factorization of

$$x_{c'}^{23}c''x_{c'}c'x_{c'}\overline{c'}x_{c'}\overline{c''}x_{c'}^{25}$$

in P', so is c'. Therefore, c' is not a complete factor in the factorization of

$$x_{c'}^{15}\overline{a_1'}x_{c'}c'x_{c'}\overline{b'}x_{c'}^{17}$$

in P'', so $\overline{a_1'}$ and $\overline{b'}$ are.

Let P''' be the prefix of P'' that is a diverse palindromic factorization of S_i'. Since $\overline{a_1'}$ and $\overline{b'}$ are complete factors later in P'', they are not complete factors in P'''. Therefore, $\overline{a_1}$ and $\bar{b}$ are complete factors in the factorizations of

$$x_{c'}^3 a_1' x_{c'} a_1 x_{c'},\ \overline{a_1} x_{c'} \overline{a_1'} x_{c'}^5 \qquad \text{and} \qquad x_{c'}^{11} b' x_{c'} b x_{c'} \bar{b} x_{c'} \overline{b'} x_{c'}^{13}$$

in P''', so they are not complete factors in the prefix $P^\dagger$ of P that is a diverse palindromic factorization of S_{i-1}'. Since we built S_{i-1}' from S_{i-1} with Lemma 2, it follows that a_1 and b are complete factors in the prefix of P that encodes τ. Therefore, τ makes the outputs of C_{i-1} labelled a and b true.

Going through all the possibilities for how P can end, which we will do in the full version of this paper, we find that each diverse palindromic factorization of S_i encodes some assignment to the inputs of C_i. This gives us the following lemma:

Lemma 3. *If we have a string S_{i-1} that represents C_{i-1} and C_i is obtained from C_{i-1} by making two outputs of C_{i-1} the inputs of a new NAND gate, then in constant time we can append symbols to S_{i-1} to obtain a string S_i that represents C_i.*

6 Conclusion

By Lemmas 1, 2 and 3 and induction, given a Boolean circuit C composed only of splitters and NAND gates with two inputs and one output, in time linear in the size of C we can build, inductively and in turn, a sequence of strings $S_1, \ldots, S_t$ such that S_i represents C_i. As mentioned in Section 2, once we have S_t we can easily build in constant time a string S that has a diverse palindromic factorization if and only if C is satisfiable. Therefore, diverse palindromic factorization is NP-hard. Since it is obviously in NP, we have the following theorem:

Theorem 1. *Diverse palindromic factorization is NP-complete.*

Acknowledgments. Many thanks to Gabriele Fici for his comments on a draft of this paper, and to the anonymous referee who pointed out a gap in the proof of Lemma 3.

References

1. Alitabbi, A., Iliopoulos, C.S., Rahman, M.S.: Maximal palindromic factorization. In: Proceedings of the Prague Stringology Conference (PSC), pp. 70–77 (2013)

2. Fernau, H., Manea, F., Mercaş, R., Schmid, M.L.: Pattern matching with variables: fast algorithms and new hardness results. In: Proceedings of the 32nd Symposium on Theoretical Aspects of Computer Science (STACS), pp. 302–315 (2015)
3. Fici, G., Gagie, T., Kärkkäinen, J., Kempa, D.: A subquadratic algorithm for minimum palindromic factorization. Journal of Discrete Algorithms **28**, 41–48 (2014)
4. Frid, A.E., Puzynina, S., Zamboni, L.: On palindromic factorization of words. Advances in Applied Mathematics **50**(5), 737–748 (2013)
5. Gawrychowski, P., Uznański, P.: Tight tradeoffs for approximating palindromes in streams. Technical Report 1410.6433, arxiv.org (2014)
6. I, T., Sugimoto, S., Inenaga, S., Bannai, H., Takeda, M.: Computing palindromic factorizations and palindromic covers on-line. In: Kulikov, A.S., Kuznetsov, S.O., Pevzner, P. (eds.) CPM 2014. LNCS, vol. 8486, pp. 150–161. Springer, Heidelberg (2014)
7. Kosolobov, D., Rubinchik, M., Shur, A.M.: Pal^{k} is linear recognizable online. In: Italiano, G.F., Margaria-Steffen, T., Pokorný, J., Quisquater, J.-J., Wattenhofer, R. (eds.) SOFSEM 2015-Testing. LNCS, vol. 8939, pp. 289–301. Springer, Heidelberg (2015)
8. Ravsky, O.: On the palindromic decomposition of binary words. Journal of Automata, Languages and Combinatorics **8**(1), 75–83 (2003)
9. Tseitin, G.S.: On the complexity of derivation in propositional calculus. In: Slisenko, A.O. (ed.) Structures in Constructive Mathematics and Mathematical Logic, Part II, pp. 115–125 (1968)
10. Ziv, J., Lempel, A.: A universal algorithm for sequential data compression. IEEE Transactions on Information Theory **22**(3), 337–343 (1977)
11. Ziv, J., Lempel, A.: Compression of individual sequences via variable-rate coding. IEEE Transactions on Information Theory **24**(5), 530–536 (1978)