

# An Infinite de Bruijn Sequence

December 25, 2016

## 1 A cycle joining Construction

**Definition 1.** A word in  $\mathbb{N}^*$  is called a *key word* if it is bigger in right-to-left lexicographical order than all of its rotations.

**Definition 2** (Cycle Construction). For  $n \in \mathbb{N}$ , let  $k_1, k_2, \dots$  be an enumeration in right-to-left lexicographic order of all the key words of length  $n$  and let  $z(k_i)$  be the number of leading zeros in  $k_i$ . Define  $C_0 = \langle 0^n \rangle$  and, for  $i > 0$ , let  $C_i$  be the sequence obtained from  $C_{i-1}$  by adding

$$\langle \text{RotLft}(k_i, z(k_i) + 1), \text{RotLft}(k_i, z(k_i) + 2), \dots, \text{RotLft}(k_i, z(k_i)) \rangle$$

after the word  $(\sigma - 1)w$  where  $\sigma w = \text{RotLft}(k_i, z(k_i))$ .<sup>1</sup>

**Proposition 3.** For every  $l > 0$  there is some  $i > 0$  such that the first  $l$  words in  $C_i$  are also the first  $l$  words in  $C_j$  for any  $j > i$ .

**Definition 4.** Let  $C_\infty$  be the infinite sequence whose prefixes are all prefixes of an infinite number of elements in  $\{C_0, C_1, \dots\}$ .

**Proposition 5.** The sequence  $C_\infty$  is a de Bruijn sequence of order  $n$ , i.e., it is a Hamiltonian path over the infinite de Bruijn graph of order  $n$  whose vertexes are the words  $\mathbb{N}^n$  and whose edges are  $\{\langle \sigma w, w \sigma' \rangle : \sigma, \sigma' \in \mathbb{N}, w \in \mathbb{N}^{n-1}\}$ .

**Definition 6** (Last on Necklace). Let  $\text{last}(w)$  be true if and only if  $w = \text{RotLft}(k_i, z(k_i))$  for some  $i$ . i.e., if  $w$  is the last member of its necklace (set of rotations) in the sequence.

**Definition 7.** Let  $\text{next}: \mathbb{N}^n \rightarrow \mathbb{N}^n$  be defined by

$$\text{next}(\sigma w) = \begin{cases} w(\sigma + 1) & \text{if } \text{last}((\sigma + 1)w) \\ w0 & \text{if } \text{last}(\sigma w) \\ w\sigma & \text{otherwise} \end{cases}$$

**Proposition 8.** For every  $w \in \mathbb{N}^n$ ,  $\text{next}(w)$  is the word that follows  $w$  in the sequence  $C_\infty$ .

---

<sup>1</sup>This is always possible because  $(\sigma - 1)w$  is a rotation of a key word that is smaller than  $k_i$  in left-to-right lexicographic order, i.e., it came earlier in the enumeration.

**Definition 9.** Let  $prv: \mathbb{N}^n \rightarrow \mathbb{N}^n$  be defined by

$$prv(w\sigma) = \begin{cases} (\sigma - 1)w & \text{if } last(\sigma w) \\ \sigma'w & \text{if } \sigma = 0 \text{ and } \sigma' \text{ is the maximal such that } last(\sigma'w) \\ \sigma w & \text{otherwise} \end{cases}$$

**Proposition 10.** For every  $w \in \mathbb{N}^n \setminus \{0^n\}$ ,  $prv(w)$  is the word that precedes  $w$  in the sequence  $C_\infty$ .

**Proposition 11.** Both  $prv$  and  $nxt$  can be computed in  $O(n^2)$  time and space.

**Proposition 12** (Prefer Maximum). For every  $\sigma \in \mathbb{N}$  and  $w \in \mathbb{N}^{n-1}$ , the word  $\sigma w$  comes before the word  $(\sigma + 1)w$  in  $C_\infty$ .

## 2 Forward and backwards transformations

**Definition 13.** For a parameter  $n$ , Let  $L \subset \mathbb{N}^+$  be the set of non-periodic words over the alphabet  $\Sigma = \mathbb{N}$  that are bigger in Arabic (right-to-left) lexicographical order than all of their rotations. Let  $L_n$  be the set of all the words in  $L$  whose length divides  $n$ .

**Definition 14.** For a word  $w = w_1 \cdots w_{n-1}w_n$  let  $R(w) = w_n w_1 \cdots w_{n-1}$  be the rotation of  $w$  to the right. Then, the nested invocation  $R^m(w)$  is the  $m$  letter rotation to the right and its inverse  $R^{-m}(w)$  is the  $m$  letter rotation to the left.

**Definition 15.** For a word  $w$  whose length is smaller or equal than  $n$ , let  $f(w)$  be the transformation defined by successive applications of the following steps to  $w$ :

- $f_1$ : Increase the first letter of the word by one.
- $f_2$ : Pad with zeros on the left to get a word of length  $n$ .
- $f_3$ : Apply the substitution rules  $u(vu)^+ \mapsto vu$  and then  $w^+ \mapsto w$ , with the longest possible  $u$  and the shortest possible  $w$ .

**Definition 16.** For a word  $w$  whose length is smaller or equal than  $n$ , let  $b(w)$  be the transformation defined by successive applications of the following steps to  $w$ :

- $b_1$ : Expand  $w$  to  $uw^m$  where  $m = \lfloor n/|w| \rfloor$  and  $u$  is the suffix of length  $n - m|w|$  of  $w$ .
- $b_2$ : Remove leading zeros.
- $b_3$ : Decrease the first letter by one.

**Observation 17.** For any  $w \in L_n$ ,  $f(b(w)) = b(f(w)) = w$ .

**Proposition 18.** *If we start with  $w(0) = 0$  and generate a sequence of words by  $w(i+1) = f(w(i))$ , we get an enumeration of all the words in  $L$  whose length is smaller or equal to  $n$ .*

*Proof.* This is a version of Duval's algorithm with a reversed order of the alphabet and a reversed order of letters in a word.  $\square$

**Definition 19.** Let  $f^*(w)$  be the first word in  $f(w), f(f(w)), \dots$  whose length divides  $n$  and, similarly, let  $b^*(w)$  be the first word in  $b(w), b(b(w)), \dots$  whose length divides  $n$ .

**Definition 20.** Let  $w(0), w(1), \dots$  be the sequence generated by starting with  $w(0) = 0$  and then continuing ad infinitum by  $w(i+1) = f^*(w(i))$  and let  $w^\infty \in \mathbb{N}^\omega$  be the concatenation of all these words.

### 3 Where can I find $w$ as a sub-word of $w^\infty$ ?

In this section we point at the position of an arbitrary word  $w$  as a sub-word of  $w^\infty$  relative to the position of the a corresponding word in  $L_n$ . This is given in Proposition 24 and in Proposition 25. Towards the proofs of these propositions, we first establish some technical results about the functions  $b$  and  $b^*$  specified, respectively, in Definition 16 and in Definition 19.

**Proposition 21.** *If  $w \in L_n$  and  $|w| \neq n$  then  $b(w) = uw$  for some non-empty word  $u$ .*

*Proof.* The first transformation  $b_1$  extends  $w$  to the left producing the word  $b_1(w) = uw^m$  where  $u$  is a tail of  $w$ . Since  $w \in L_n$  and because it contains a letter  $\sigma$  that is not zero, we have, by maximality of  $w$  among its rotations in right-to-left lexicographical order, that its last letter is not zero. The last letter of  $u$  is the last letter of  $w$  so it is also not zero. This gives us that the next transformation  $b_2$ , that deletes trailing zeros, leaves at least the last copy of  $w$  and the last letter of the before-last (full or partial) copy at the tail of  $b_1(w)$ . Thus,  $b_2(b_1(w)) = uw$  where  $u$  is a non-empty word whose first letter is not zero. Then, the last transformation  $b_3$  only decreases the first letter of  $u$  by one which gives us that  $b(w) = b_3(b_2(b_1(w))) = vw$  for some non-empty word  $v$ .  $\square$

**Proposition 22.** *For any  $w = 0^l \sigma \hat{w} \in L_n$  where  $\sigma$  is a non-zero letter there is a non-empty word  $u$  such that  $b(w) = u\hat{w}$ .*

*Proof.* If  $|w| \neq n$  the proof follows by Proposition 21. If  $|w| = n$  then  $b_1(w) = w$ ,  $b_2(b_1(w)) = \sigma \hat{w}$ , and  $b_3(b_2(b_1(w))) = (\sigma - 1)\hat{w}$  and the claim follows as well.  $\square$

**Proposition 23.** *Let  $w$  be an arbitrary word in  $\mathbb{N}^n$  and let  $\bar{w} = f_3(w)$ . Let  $l$  be the (possibly zero) number of trailing zeros (from the left) in  $w$ . Then, for all  $0 \leq i \leq |w| - l - 1$ , the word  $R^i(\bar{w})$  comes  $i + n - |w|$  letters before  $w$  as a sub-word of  $w^\infty$ .*

**Proposition 24.** For a given  $w \in L_n$ , let  $l$  be the number of trailing zeros (from the left) in  $w$  and let  $\bar{w} = b_1(w)$ . Then, for all  $0 \leq i \leq |w| - l - 1$ , the word  $R^i(\bar{w})$  comes  $i + n - |w|$  letters before  $w$  as a sub-word of  $w^\infty$ .

*Proof.* By Proposition ?? the words that come before  $w$  ends with the last  $|w| - l$  letters of  $w$ . In particular, the  $n$  letter word that starts  $i + n - |w|$  before  $w$  is  $R^i(\bar{w})$ .  $\square$

**Proposition 25.** For a given  $w \in L_n$ , let  $l$  be the number of trailing zeros (from the left) in  $w$  and let  $\bar{w} = b_1(w)$ . Then, for all  $|w| - l \leq i \leq n - 1$  the word  $R^i(\bar{w})$  comes  $i - (n - |f_3(u)| \pmod{n})$  letters before the first  $u \in \langle 0^{m-1}(\bar{w}_m + 1)\bar{w}_{m+1} \cdots \bar{w}_n \rangle_{m=i+1}^n$  that is in  $L_n$ .

**Proposition 26.** The word  $w^\infty$  contains all the words in  $\mathbb{N}^n$  as subwords.

*Proof.* Any word of length  $n$  is a rotation of the expansion of a word in  $L_n$ .  $\square$

**Proposition 27.** For any  $k$  the prefix  $w_1^\infty \cdots w_{k^n}^\infty$  is an  $n$ -order de Bruijn sequences. Moreover, it is the reversed of the  $n$ -order prefer-max sequence on the alphabet  $\langle 0, \dots, k-1 \rangle$  (in this order).

*Proof.* Counting argument + arguing that if  $|w| = n - 1$  and  $\sigma_1 < \sigma_2$  then  $w\sigma_1$  comes before  $w\sigma_2$  as subwords of  $w^\infty$ .  $\square$

**Proposition 28.** For  $w \in \mathbb{N}^n$ , let  $i$  be the minimal index such that  $R^{-i}(w) \in L$  and let  $\bar{w} = R^{-i}(w)$ . Let  $\bar{w}^+ = \bar{w}_{1..i}(\bar{w}_{i+1} + 1)\bar{w}_{(i+2)..n}$ , i.e., the word obtained by increasing the  $(i + 1)$ th letter of  $\bar{w}$  by one. Then, the function

$$\text{next}(w) = \begin{cases} f^*(f_3(w))_1 & \text{if } w \in L; \\ w_1 + 1 & \text{if } \bar{w}_{1..i} = 0^i \wedge \bar{w}^+ \in L \wedge \max(\bar{w}_{1..(n-1)}^+) \leq \max(w); \\ 0 & \text{if } \bar{w}_{1..i} = 0^i \wedge (\bar{w}^+ \notin L \vee \max(\bar{w}_{1..(n-1)}^+) > \max(w)); \\ w_1 & \text{otherwise.} \end{cases}$$

represents the mapping of a word  $w$  to the letter that follows the (one and only) occurrence of  $w$  as a subword of  $w^\infty$ .

**Definition 29.** Let  $w(0) = 0, w(1) = f^*(w(0)), \dots, w(i) = f^*(w(i-1)), \dots$  be our enumeration of all the words in  $L_n$ . Let  $w^{(i)} = w(0) \cdots w(i)$  be the concatenation of the first  $i$  words in this enumeration and let  $u(j) = w_{j-n+1}^{(i)} \cdots w_j^{(i)}$  be the “window” of length  $n$  before the  $j$ th letter in  $w^{(i)}$ .

**Proposition 30.** Let  $w = w(i)$  for some  $i$  and let  $l$  be the number of leading zeros in  $w$ . Then, inserting the cycle  $\langle R^{-l-n-1}(w), \dots, R^{-l}(w) \rangle$  to  $\langle u(j) \rangle_{j=0}^{i-1}$  after the word obtained from  $R^{-l}(w)$  by decreasing its first letter by one yields the sequence  $\langle u(j) \rangle_{j=0}^i$ .

## 4 Where can I find $w$ as a sub-word of $w^\infty$ ? (second try...)

**Definition 31.** For a word  $w$ ,  $\max(w)$  is the maximal digit in  $w$ .

**Definition 32.** A word  $u \in \mathbb{N}^n$  corresponds to  $w \in L_n$  if  $u$  is a rotation of  $w^{\lfloor \frac{n}{|w|} \rfloor}$ . Note that each  $u \in \mathbb{N}^n$  corresponds to exactly one word  $w \in L_n$ .

**Proposition 33.** If  $w \in L_n$  and  $|w| < n$ , then  $f^*(w) = f(w) = 0^{n-|w|}x$  for some word  $x$ .

*Proof.* Write  $f_1(w) = x$ ,  $f_2(x) = 0^{n-|w|}x$ . Since  $w \in L_n$  and  $|w| < n$ ,  $n - |w| \geq \frac{n}{2}$ . Moreover, the last digit in  $x$  is not zero. Hence,  $f(w) = f_3(0^{n-|w|}x) = 0^{n-|w|}x$ . Since  $|0^{n-|w|}x| = n$ , we have  $f(w) = f^*(w) = 0^{n-|w|}x$ .  $\square$

**Proposition 34.** Take  $|w| < n$  so that  $w = w'k$  where  $0 < k = \max(w)$ , then  $b(w) = uw$  and  $\max(u) \leq \max(w)$ .

*Proof.* Write  $w = w'k$ . Thus,  $b_1(w) = xk(w'k)^r$ ,  $r > 0$ .  $b_2(xk(w'k)^r) = y(w'k)^r$ .  $b_3(y(w'k)^r) = uw'k = uw$ . It is easy to see that  $\max u \leq k$ .  $\square$

**Proposition 35.** If  $w \in L_n$ ,  $|w|^m = n$ ,  $m > 1$  and  $w = 0^l \sigma \hat{w}$  such that  $\sigma \neq 0$ , then  $b^*(w) = u\hat{w}w^{m-1}$  for some  $u$ .

*Proof.* Since  $w \in L_n$  and  $w \neq 0$ ,  $b(w)$  is defined and

$$b(w) = (\sigma - 1)\hat{w}w^{m-1}.$$

If  $b(w) \in L_n$  we are done, and otherwise  $|b(w)| < n$  and several invocations of the previous proposition provide the required.  $\square$

**Proposition 36.** Assume that  $u \in \mathbb{N}^n$  corresponds to  $w \in L_n$  such that  $|w| < n$ . then,  $u$  is a subword of  $w^\infty$ .

*Proof.* If  $u = 0^n$ , then  $u$  is a prefix of  $w^\infty$  and we are done. Otherwise,  $w = 0^l \sigma \hat{w}$  where  $\sigma \neq 0$ . Take  $m$  such that  $|w|^m = n$ . Note that  $m > 1$ . By Propositions 33 and 35,  $b^*(w)wf^*(w) = x\hat{w}w^{m-1}w0^{|w|}y$ , which is also a subword of  $w^\infty$ . Hence,

$$\hat{w}(0^l \sigma \hat{w})^m 0^l \text{ is a subword of } w^\infty.$$

$u$  is a rotation of  $w^m$  thus  $u$  is a subword of  $\hat{w}(0^l \sigma \hat{w})^m 0^l$  which implies that  $u$  is a subword of  $w^\infty$ .  $\square$

**Proposition 37.** Assume that  $u = yx \in \mathbb{N}^n$  corresponds to  $w = xy \in L_n$  where  $|w| = n$ . If  $x \neq 0^r$ , then  $u$  is a subword of  $w^\infty$ .

*Proof.* We show that  $u = yx$  is a subword of  $b^*(w)w$ . Write  $x = 0^l \sigma z$  where  $\sigma \neq 0$ . Thus, since  $|w| = n$ ,  $b(w) = (\sigma - 1)zy$ . If  $b(w) = b^*(w)$ , then

$$b^*(w)w = (\sigma - 1)zyx$$

and we get that  $u$  is a subword of  $b^*(w)b(w)$ . Otherwise,  $|(\sigma - 1)zy|$  does not divide  $n$ , and in particular,  $|(\sigma - 1)zy| < n$ . By applying Proposition 34 several times, we get that  $b^*(w) = v(\sigma - 1)zy$  for some  $v$ , and  $u = yx$  is a subword of  $b^*(w)w = v(\sigma - 1)zyxyx$ .  $\square$

**Lemma 38.** Assume that  $w = 0^l v \in L_n$  and  $|w| = n$ . Write  $w = 0^l z_1 \sigma z_2$  where  $\sigma$  is the first digit in  $v$  such that  $0^{l+|z_1|}(\sigma + 1)z_2$  is lexicographically maximal among its rotations. Take  $k \in \mathbb{N}$  and a suffix of  $(\sigma z_2)$ ,  $u$  such that  $|u(\sigma z_2)^{k+1}| = |z_1(\sigma z_2)|$ . Then,  $u(\sigma z_2)^{k+1} = z_1(\sigma z_2)$ .

*Proof.* Assume for a contradiction that the claim is false, and hence  $z_1 \neq u(\sigma z_2)^k$ . Therefore, there are  $\tau \neq \tau'$  in  $\mathbb{N}$  and a word  $y$ , such that  $\tau'y$  is a suffix of  $\sigma z_2$ , and

$$z_1 = x\tau y(\sigma z_2)^r, \quad (\sigma z_2)^k = x'\tau'y(\sigma z_2)^r.$$

Clearly,  $\tau < \tau'$  since otherwise,  $\tau' < \tau$ , and we get that  $w = 0^l z_1 \sigma z_2 = 0^l x\tau y(\sigma z_2)^{r+1}$ . However, if we assume that  $\tau' < \tau$ ,  $w' = (\sigma z_2)^r 0^l x\tau'y$  is lexicographically larger than  $w$ , in contradiction to  $w \in L_n$ .  $\square$

**Corollary 39.** Assume that  $w = 0^l v \in L_n$  and  $|w| = n$ . Write  $w = 0^l z_1 \sigma z_2$  where  $\sigma$  is the first digit in  $v$  such that  $0^{l+|z_1|}(\sigma + 1)z_2$  is lexicographically maximal among its rotations. Then, there are words  $x, y$  such that  $z_2 = xy$ ,  $w = 0^l y(\sigma xy)^{r+1}$  and  $z_1 = y(\sigma xy)^r$ .

*Proof.* This is a consequence of the previous Lemma and the fact that  $|0^l z_1| = |x(\sigma z_2)|^m$ .  $\square$

**Lemma 40.** If  $uv = vu$  and  $u, v \neq \varepsilon$ , then there is some word  $w$ , such that  $u, v \in \{w\}^*$ .

*Proof.* By induction on  $|u| + |v|$ . If  $|u| = |v|$ ,  $u = v$  and we are done. Otherwise, assume w.l.o.g. that  $|u| > |v|$  and write  $u = vx$  (since  $uv = vu$ ). Then,  $ux = vxv = vvx = vu$ . We see that  $xv = vx$ . By the induction hypothesis,  $x = w^k$  and  $v = w^l$ . Hence,  $u = w^{l+k}$  as required.  $\square$

**Lemma 41.** Let  $w = 0^l y(x0^l y)^{r+1}$  be an  $n$ -length word such that  $y \notin \{0\}^*$ . Then,  $w \notin L_n$ .

*Proof.* Assume for a contradiction that  $w$  is a key-word of length  $n$ , and take a maximal  $t \in \mathbb{N}$  such that  $x0^l y = x'(0^l y)^{t+1}$ . First, we note that  $x' \neq \varepsilon$ . Indeed, if  $x' = \varepsilon$ , then  $w = (0^l y)(0^l y)^{(t+1)(r+1)}$ , a periodic word, and then  $w \notin L_n$ .

Now we claim that  $|x'| < |0^l y|$ . For verifying this claim, assume that  $|x'| \geq |0^l y|$  and write  $x' = x'_1 x'_2$ , where  $|x'_2| = |0^l y|$ . By maximality of  $t$ ,  $x'_2 \neq 0^l y$ , and since  $w \in L_n$ ,  $x'_2 <_{lex} 0^l y$ . Therefore,

$$w' = (x0^l y)^r x'_1 x'_2 (0^l y)^{t+1} 0^l y$$

is a rotation of  $w$  which is lexicographically larger than  $w$ , in contradiction to  $w \in L_n$ .

To summary our conclusions, we have  $w = 0^l y(x'(0^l y)^{t+1})^{r+1} \in L_n$ , and  $|x'| < |0^l y|$ . Write  $0^l y = z_1 z_2$  where  $|x'| = |z_2|$ . Therefore,

$$w = z_1 z_2 (z_2 (z_1 z_2)^{t+1}) \dots (z_2 (z_1 z_2)^{t+1}).$$

We look now at a rotation of  $w$ ,  $w' = (z_2 (z_1 z_2)^{t+1}) \dots (z_2 (z_1 z_2)^{t+2})$ . Since  $w \in L_n$ ,  $w$  is lexicographically larger than  $w'$  and in particular,  $(z_1 z_2 z_2 (z_1 z_2)^{t+1}) \geq_{lex} (z_2 (z_1 z_2)^{t+2})$  which implies that  $z_1 z_2 z_2 \geq_{lex} z_2 z_1 z_2$ , and hence

$$z_1 z_2 \geq_{lex} z_2 z_1.$$

In addition,  $z_2 z_1 z_2$  is a suffix of  $w$  while  $z_1 z_2 z_2$  is a subword of  $w$ . Hence, as  $w \in L_n$  we have,  $z_2 z_1 z_2 \geq_{lex} z_1 z_2 z_2$ , and hence

$$z_2 z_1 \geq_{lex} z_1 z_2.$$

As a result,  $z_2 z_1 = z_1 z_2 m$  and then by Lemma 40,  $z_1 = z^{l_1}$  and  $z_2 = z^{l_2}$  for some non empty word  $z$ . Therefore,  $w = z^m$  for some  $z > 0$  in contradiction to  $w \in L_n$ . □

**Proposition 42.** Assume that  $v0^l \in \mathbb{N}^n$  corresponds to  $w = 0^l v \in L_n$  where  $|w| = n$  and  $l > 0$ . Then,  $v0^l$  is a subword of  $w^\infty$ .

*Proof.* Write  $w = 0^l z_1 \sigma z_2$  where  $\sigma \in \mathbb{N}$  is the first digit in  $w$  so that  $0^{l+|z_1|}(\sigma + 1)z_2$  is lexicographically maximal among its rotations. Note that such a digit exists since the last digit in  $w$  satisfies this requirement. Hence,  $v = z_1 \sigma z_2$ .

By Corollary 39,  $z_2 = xy$  and  $z_1 = y(\sigma xy)^r$ . Now, since  $|0^{l+|z_1|}(\sigma + 1)z_2| = n$  and  $0^{l+|z_1|}(\sigma + 1)z_2$  is lexicographically maximal among its rotations,  $0^{l+|z_1|}(\sigma + 1)z_2 = (w')^{k+1}$  where  $w' \in L_n$ . Note that  $0^{l+|z_1|}$  is a prefix of  $w'$ . We consider three possibilities

Case 1.  $\sigma z_2 \in L_n$ . We show that in this case,  $v0^l$  is a subword of  $b^*(b^*(w'))(b^*(w'))w'$ , which is a subword of  $w^\infty$ .

$b_1(w') = w'^{k+1} = 0^{l+|z_1|}(\sigma + 1)z_2$ . Hence,  $b(w') = b_3(b_2(0^{l+|z_1|}(\sigma + 1)z_1)) = \sigma z_2$ . Since  $\sigma z_2 \in L_n$ ,  $b(w') = b^*(w') = \sigma z_2$  and in particular,  $|(\sigma z_2)^{m+1}| = n$  for some  $m \in \mathbb{N}$ . Observe that  $|z_1| \leq |\sigma z_2|^m$  and use Lemma 38 to conclude that  $z_1$  is a suffix of  $(\sigma z_2)^m$ .

By invoking Proposition 34 several times,  $b^*(\sigma z_2) = u(\sigma z_2)^m$  for some  $u$ . Hence,  $v0^l = z_1 \sigma z_2 0^l$  is a subword of

$$b^*(b^*(w'))b^*(w')b(w') = u(\sigma z_2)^{m+1}0^{l+|z_1|}x'.$$

Before we deal with the other cases, we note that  $b_1(\sigma z_2) = b_1(\sigma xy) = x'y(\sigma xy)^{r+1}$  for some  $x'$  that satisfies  $|x'| = l > 0$ .

Case 2.  $\sigma z_2 \notin L_n$  and  $x' \neq 0^l$ . We show that in this case,  $v0^l$  is a subword of  $b^*(w')w'$ , which is a subword of  $w^\infty$ .

Recall that  $b(w') = \sigma z_2$  which is, by assumption, not a key-word. Since  $x' \neq 0^l$ , several invocations of Proposition 34 imply that  $b^*(\sigma z_2) = x''y(\sigma z_2)^{r+1}$ . Since  $v = z_1\sigma z_2 = y(\sigma z_2)^{r+1}$ , we get that  $v0^l$  is a subword of

$$b^*(w')w' = x''y(\sigma z_2)^{r+1}0^{l+|z_1|}u.$$

Case 3.  $\sigma z_2 \notin L_n$  and  $x' = 0^l$ . In this case,  $b_1(\sigma z_2) = 0^l y(\sigma x y)^{r+1}$ . Note that  $w = 0^l y(x''0^l y)^{r+1}$  and use Lemma 41 to obtain a contradiction.

□

**Theorem 43.**  $w^\infty$  is an infinite de Bruijn sequence.

*Proof.* According to Propositions ??, every  $n$ -sequence is a subword of  $w^\infty$ . By the “onion theorem” and by the pigeonhole principle, every  $n$ -sequence appears only once at  $w^\infty$ . □

## 5 Constructing an Infinite de Bruijn Cycle

A key word is an  $n$ -length word that is (arabic) maximal among its rotations. Let  $kw_0, kw_1, kw_2, \dots$  be an enumeration of all key-words, ordered lexicographically. Let  $C_m$  be the cycle of  $kw_m$ . We order the elements of  $C_m$  as follows: if  $kw_m = 0^l(\sigma + 1)w$ , then  $w0^l(\sigma + 1)$  is the first sequence in  $C_m$ , and each word  $w'$  is followed by  $R(w')$ . The last word in  $C_m$  is  $(\sigma + 1)w0^l$ .

For each  $m < n$  we define a de Bruijn sequence  $D_m$ , over the words  $\bigcup_{i=0}^m C_i$  as follows:

- $D_0 = 0^n$ .
- If  $kw_{m+1} = 0^l(\sigma + 1)w$ , then  $D_{m+1}$  is obtained by inserting the sequence  $C_{m+1}$  after the word  $\sigma w0^l \in C_m$ .

**Definition 44.** For  $w, w' \in \mathbb{N}^n$  and  $m \in \mathbb{N}$ , write  $w <_m w'$  if  $w$  appears before  $w'$  in  $D_m$ . Write  $< = \bigcup_{i=0}^\infty$ .

For a word  $w$ ,  $\max(w)$  is the maximal number in  $w$ .

**Lemma 45.** If  $w < w'$ , then  $\max(w) \leq \max(w')$ .

**Corollary 46.**  $<$  defines an infinite de Bruijn sequence.

*Proof.* By the previous Lemma, each word is preceded by finitely many words thus  $<$  defines an infinite sequence. Since each  $D_m$  is a de Bruijn sequence and since  $<_m \subseteq <_{m+1}$ , the sequence defined by  $<$  is a de Bruijn sequence. □

Let  $D$  be the infinite de Bruijn sequence defined by  $<$ .



**Theorem 47.** For a word  $\sigma w$  in  $D$ , let  $\text{next}(\sigma w)$  be the successor of  $\sigma w$  in  $D$ . Then,

$$\text{next}(\sigma w) = \begin{cases} w(\sigma + 1) & \text{if } \text{last}((\sigma + 1)w); \\ w0 & \text{if } \text{last}(\sigma w); \\ w\sigma & \text{otherwise} \end{cases}$$

**Definition 48.** If  $kw_m = 0^l(\sigma + 1)w$ , we write  $\text{first}(w0^l(\sigma + 1)w)$ ,  $\text{key}(0^l(\sigma + 1)w)$  and  $\text{last}((\sigma + 1)w0^l)$ . In addition, for a cycle  $C_m$ ,  $\text{first}(C_m) = w \in C_m$  so that  $\text{first}(w)$ ,  $\text{key}(C_m) = kw_m$  and  $\text{last}(C_m) = w \in C_m$  so that  $\text{last}(w)$ .

**Lemma 49.** For any  $w' \in C_m$ ,  $\text{first}(C_m) \leq w' \leq \text{last}(C_m)$ .

*Proof.* This the way we ordered the cycles.  $\square$

**Definition 50.** We say that  $C_{i_0}C_{i_1} \dots C_{i_k}$  is a sequence of cycles, if for every  $j < k$ ,  $\text{next}(\text{last}(C_{i_j})) = \text{first}(C_{i_{j+1}})$ .

**Lemma 51.** Let  $C_{i_0}C_{i_1} \dots C_{i_k}$  be a sequence of cycles. Write,  $\text{key}(C_{i_0}) = 0^l(\sigma)w$  where  $\sigma \neq 0$ . Then, for every  $j \leq k$ ,  $\text{key}(C_{i_j}) = 0^l(\sigma + j)w$ .

*Proof.* Assume by induction that  $\text{key}(C_{i_j}) = 0^l(\sigma + j)w$  for  $j < k$ . Hence,  $\text{last}(C_{i_j}) = (\sigma + j)w0^l$ . Thus,  $\text{next}((\sigma + j)w0^l) = w0^l(\sigma + j + 1)$  or  $\text{next}((\sigma + j)w0^l) = w0^{l+1}$ . Since  $\neg(\text{first}(w0^{l+1}))$ , we have

$$\text{next}((\sigma + j)w0^l) = w0^l(\sigma + j + 1) = \text{first}(C_{i_{j+1}}).$$

We conclude that  $\text{key}(C_{i_{j+1}}) = 0^l(\sigma + j + 1)w$ .  $\square$

**Lemma 52** (The parentheses property). For any two cycles  $C_k$  and  $C_m$ , one of the following occur

- $\text{last}(C_k) < \text{first}(C_m)$  or  $\text{last}(C_m) < \text{first}(C_k)$ .
- $\text{first}(C_k) < \text{first}(C_m) \leq \text{last}(C_m) < \text{last}(C_k)$  or  $\text{first}(C_m) < \text{first}(C_k) \leq \text{last}(C_k) < \text{last}(C_m)$ .

*Proof.* This concluded by the way  $D_{m+1}$  is obtained from  $D_m$ .  $\square$

**Definition 53.** We say that  $C_m$  is embedded in  $C_k$ , if  $C_m = C_k$  or  $\text{first}(C_k) < \text{first}(C_m) \leq \text{last}(C_m) < \text{last}(C_k)$ .

In addition,  $C_m$  is said to be immediately embedded in  $C_k$  if there is no  $C_l$  such that  $C_m$  is embedded in  $C_l$  and  $C_l$  is embedded in  $C_k$ .

We define by inductively the statement: “ $C_m$  is  $r$ -embedded in  $C_k$ ”:

- $C_m$  is 0-embedded in  $C_k$  if  $C_m = C_k$ .
- $C_m$  is  $r$ -embedded in  $C_k$  if there is a cycle  $C_l$  such that  $C_m$  is  $r - 1$ -embedded in  $C_l$  and  $C_l$  is immediately embedded in  $C_k$ .

**Lemma 54.** Assume that  $C_m$  is immediately embedded  $C_k$ . Write  $\text{key}(C_m) = 0^i(\sigma + 1)0^jw$  where  $w$  does not starts with 0. Then,

- $key(C_k) = 0^{i+1+j}w$ .
- If  $u \in C_k$  and  $last(C_m) < u$ , then  $u = 0^{j_2}w0^{i+1+j_1}$  where  $j_1 + j_2 = j$ .

*Proof.* To prove the first item, note that since  $0^i(\sigma+1)0^jw$  is maximal among its rotations, the same holds for  $0^{i+1+j}w$ . Thus, we need to show that  $0^{i+1+j}w \in C_k$ .

Take a maximal sequence of cycles that begins in  $C_m$  and let  $C_{m+r}$  be the last cycle in this sequence. By lemma 51,  $key(C_{m+r}) = 0^i(\sigma+1+r)0^jw$ . By the parentheses property,  $C_{m+r}$  is immediately embedded in  $C_k$ . And finally, by the maximality of the sequence of cycles,  $next(last(C_{m+r})) \in C_k$ .

As a result, we have:

$$next(last(C_{m+r})) = next((\sigma+1+r)0^jw0^i) \in \{0^jw0^i0, 0^jw0^i(\sigma+2+r)\}.$$

If  $next((\sigma+1+r)0^jw0^i) = 0^jw0^i(\sigma+2+r)$ , then  $last((\sigma+2+r)0^jw0^i)$  which implies  $first(0^jw0^i(\sigma+2+r))$ . But  $first(next(last(C_{m+r})))$  contradicts the maximality of our sequence of cycles. Hence,  $next((\sigma+1+r)0^jw0^i) = 0^jw0^i0 \in C_k$ . Now,  $0^{i+1+j}w$  is a rotation of  $0^jw0^i0$  thus  $0^{i+1+j}w \in C_k$  as required.

For proving the second item, we note that  $last(C_k) = w0^{i+1+j}$ . As we have seen, the first element in  $C_k$  that follows  $C_m$  (and follows  $C_{m+r}$ ) is  $next(last(C_{m+r})) = 0^jw0^i0$ . As a result, if  $u \in C_k$  and  $last(C_m) < u$ , then

$$0^jw0^i0 \leq u \leq w0^{i+1+j}$$

which implies that  $u = 0^{j_2}w0^{i+1+j_1}$  □

**Corrolary 55.** Assume that  $C_m$  is  $r$ -embedded in  $C_k$ . Write  $key(C_m) = uv$  where  $u$  is the minimal prefix of  $key(C_m)$  that includes  $r$  non-zero numbers. Then,  $key(C_k) = 0^{|u|}v$ .

*Proof.* By  $r$  invocations of the first item of the previous lemma. □

**Lemma 56.** Assume that  $last(C_k) < first(C_m)$  and  $C_k, C_m$  are both immediately embedded in a cycle  $C$ . Write  $key(C_k) = 0^lw$  where  $w$  starts with a non-zero letter. Then,  $key(C_m) = 0^lw'$  where  $w <_{lex} w'$ .

*Proof.* First, if there is a sequence of cycles from  $C_k$  and  $C_m$ , the claim follows from Lemma 51. Otherwise, there is some  $u \in C$  such that

$$last(C_k) < u < first(C_m).$$

Consider a maximal sequence of cycles that starts with  $C_k$ . This sequence ends in some cycle  $C_{k'}$ . Similarly, consider a maximal sequence of cycles that ends in  $C_m$  and let  $C_{m'}$  be the first element in this sequence. Therefore,

$$last(C_{k'}) < v < first(C_{m'}) \implies v \in C.$$

Write  $key(C_k) = 0^l(\sigma+1)w_1$  (namely,  $w = (\sigma+1)w_1$ ) and write  $w_1 = 0^jw_2$  where  $w_2$  starts with non-zero letter. Hence,

$$key(C_k) = 0^l(\sigma+1)0^jw_2. \tag{1}$$

By Lemma 51,  $key(C_{k'}) = 0^l(\sigma + 1 + r)0^j w_2$ , and hence  $last(C_{k'}) = (\sigma + 1 + r)0^j w_2 0^l$ . Let  $u_1$  be the successor of  $last(C_{k'})$ . Thus,

$$u_1 = 0^j w_2 0^{l+1}.$$

Now, let  $u_2$  be the predecessor of  $first(C_{m'})$ . Thus,

$$u_2 = 0^{j_2} w_2 0^{l+1+j_1}, \text{ where } j_1 + j_2 = j.$$

In addition, since  $C_{m'}$  is embedded in  $C$ ,  $u_2 \neq last(C)$  thus  $j_2 > 0$ . As a result,

$$first(C_{m'}) = 0^{j_2-1} w_2 0^{l+1+j_1} 1.$$

We get that  $key(C_{m'}) = 0^{l+1+j_1} 1 0^{j_2-1} w_2$ . Hence, by Lemma 51,

$$key(C_m) = 0^{l+1+j_1} (1+t) 0^{j_2-1} w_2. \quad (2)$$

By Equation 1,  $w = (\sigma + 1) 0^j w_2$ . By Equation 2,  $w' = 0^{l+1+j_1} (1+t) 0^{j_2-1} w_2$ . We see that indeed  $w <_{lex} w'$ .  $\square$

**Lemma 57.** *If  $k < m$ , then  $first(C_k) < first(C_m)$ .*

*Proof.* Since  $<$  is a linear ordering, we can prove an equivalent statement:

$$first(C_k) < first(C_m) \implies k < m.$$

We take such cycles  $C_k$  and  $C_m$ . By the parentheses property, either  $C_m$  is embedded in  $C_k$ , or  $C_m$  is entirely after  $C_k$ . If  $C_m$  is embedded in  $C_k$ , by Corollary 55 we get that  $k < m$ . It is left to deal with the case that  $last(C_k) < first(C_m)$ . We consider two cases.

**Case 1.**  $C_k$  and  $C_m$  are both embedded in some cycle  $C$ .

In this case, we can find cycles  $C_{k'}$  and  $C_{m'}$  such that

1.  $C_k$  is embedded in  $C_{k'}$  and  $C_m$  is embedded in  $C_{m'}$
2.  $C_{k'}$  and  $C_{m'}$  are immediately embedded in  $C$ .

By item 1 and Lemma 55, we can write  $key(C_k) = uv$  and  $key(C_{k'}) = 0^{|u|}v$ . Write

$$v = 0^l v_1$$

where  $v_1$  starts with a non-zero letter thus  $key(C_{k'}) = 0^{|u|+l} v_1$ . By item 2 and Lemma 56,  $key(C_{m'}) = 0^{|u|+l} v_2$  where

$$v_1 <_{lex} v_2.$$

Write  $v_2 = 0^r v'_2$  where  $v'_2$  starts with a non zero letter. Hence,

$$v_1 = x v'_1$$

where  $|x| = r$ ,  $|v'_1| = |v'_2|$ , and

$$v'_1 <_{lex} v'_2.$$

Since  $key(C_{m'}) = 0^{|u|+l}v_2 = 0^{|u|+l+r}v'_2$ , by item 2 and Lemma 55,  $key(C_m) = u'v'_2$ . Recall that  $key(C_k) = uv = u0^lv_1 = u0^lxv'_1$ . Since  $v'_1 <_{lex} v'_2$ ,  $key(C_k) <_{lex} key(C_m)$  thus  $k < m$  as required.

**case 2.** There is no cycle  $C$  such that  $C_k$  and  $C_m$  are embedded in  $C$ .

Take a cycle  $C_{k'}$  such that  $C_k$  is  $r_1$ -embedded in  $C_{k'}$  and  $r_1$  is maximal with respect to this property. Similarly, take a cycle  $C_{m'}$  such that  $C_m$  is  $r_2$ -embedded in  $C_{m'}$  and  $r_2$  is maximal as possible. By the parentheses property,  $C_{k'}$  is entirely before  $C_{m'}$ .  $C_{k'}$  and  $C_{m'}$  are not embedded in any cycle so there is a sequence of cycles from  $C_{k'}$  to  $C_{m'}$ . Write  $key(C_k) = u(\sigma + 1)v$  such that  $key(C_{k'}) = 0^{|u|}(\sigma + 1)v$  (by Lemma 55). Hence,  $key(C_{m'}) = 0^{|u|}(\sigma + 1 + r)v$  where  $r > 1$  (by Lemma 51). Finally, by Lemma 55 we get  $key(C_m) = u'(\sigma + 1 + r)v$ . We see that  $key(C_k) = u(\sigma + 1)v <_{lex} u'(\sigma + 1 + r)v = key(C_m)$ . Therefore,  $k < m$ .

□

**Theorem 58.** For any word  $\tau w$ ,  $\tau w < (\tau + 1)w$ .

*Proof.* Fix an arbitrary cycle  $C_m$ . We show that if  $(\tau + 1)w \in C_m$ . Then,  $\tau w < (\tau + 1)w$ . We start by showing this fact for  $last(C_m)$ . Write  $key(C_m) = 0^l(\sigma + 1)w'$  and hence,  $last(C_m) = (\sigma + 1)w'0^l$  and  $first(C_m) = w'0^l(\sigma + 1)$ . Write  $pre = nxt^{-1}$  and note that  $pre(w'0^l(\sigma + 1)) = \sigma w'0^l$ . Hence, we get that

$$\sigma w'0^l < (\sigma + 1)w'0^l.$$

Now we deal with the general case in which  $(\tau + 1)w \neq last(C_m)$ . We write  $key(C_m) = 0^l(\sigma + 1)w_1(\tau + 1)w_2$ , where

$$(\tau + 1)w = (\tau + 1)w_20^l(\sigma + 1)w_1.$$

Let  $C_k$  be the cycle of  $\tau w$ . Since every rotation of  $\tau w$  is lexicographically smaller than some rotation of  $(\tau + 1)w$ , we have  $key(C_k) <_{lex} key(C_m)$ . Hence,  $k < m$  and by Lemma 57, we get

$$first(C_k) < first(C_m).$$

Now, if  $last(C_k) < first(C_m)$  then every element of  $C_k$  precedes every element of  $C_m$  and we are done. Otherwise, by the parentheses property,  $C_m$  is embedded in  $C_k$ . Note that  $|(\tau + 1)w|_0 - |\tau w|_0 \in \{0, 1\}$ . Use corollary 55 to conclude that  $|(\tau + 1)w|_0 - |\tau w|_0 = 1$  and that  $C_m$  is immediately embedded in  $C_k$ . Moreover, note that as  $|(\tau + 1)w|_0 - |\tau w|_0 = 1$ , we have  $\tau = 0$ .

According to our conclusions, we can write:  $key(C_m) = 0^l(\sigma + 1)w_11w_2$ . We write  $w_1 = 0^jw'_10^i$  and  $w_2 = 0^rw'_2$  where  $w'_1$  and  $w'_2$  do not start or end with zero. We have,

$$key(C_m) = 0^l(\sigma + 1)0^jw'_10^i10^rw'_2.$$

Assume for a contradiction that  $(\tau + 1)w < \tau w$ . Recall that  $\tau w \in C_k$  and that  $C_m$  is immediately embedded in  $C_k$ , and conclude that  $last(C_m) < \tau w \leq last(C_k)$ . Therefore, by Lemma 54,  $\tau w = 0^{j_2}w'_10^i10^rw'_20^{l+1+j_1}$ . Thus, we have:

$$0^{j_2}w'_10^i10^rw'_20^{l+1+j_1} = 0^{r+1}w'_20^l(\sigma + 1)0^jw'_10^i. \quad (3)$$

Since also  $|0^{j_2}w'_10^i10^rw'_20^{l+1+j_1+1}|_1 = |0^{r+1}w'_20^l(\sigma+1)0^jw'_10^i|_1$  we get  $\sigma+1 = 1$ . Hence,

$$key(C_m) = 0^l10^jw'_10^i10^rw'_2 \quad (4)$$

and Equation 3 can be rewritten as follows:

$$0^{j_2}w'_10^i10^rw'_20^{l+1+j_1} = 0^{r+1}w'_20^l10^jw'_10^i. \quad (5)$$

For the rest of the proof we assume that  $w'_1 \neq \varepsilon$  and  $w'_2 \neq \varepsilon$ . The other cases are dealt similarly<sup>2</sup>. By deleting the initial and final segments of zeroes, we get from Equation 5,

$$j_2 = r + 1, \quad w'_10^i10^rw'_2 = w'_20^l10^jw'_1. \quad (6)$$

Now, by Equation 4,

$$0^l10^jw'_10^i10^rw'_2 \geq_{lex} 0^i10^rw'_20^l10^jw'_1. \quad (7)$$

By combining equations 6 and 7, we get

$$0^l10^j \geq_{lex} 0^i10^r. \quad (8)$$

Hence,  $j \leq r$  and in particular,  $j_2 \leq r$  in contradiction to Equation 6.  $\square$

---

<sup>2</sup>really! I checked !