Title
REPEAT-FREE SEQUENCES

Author
Elser, V.

Publication Date
1983-09-01

LBL-16632
UC-32 c.2

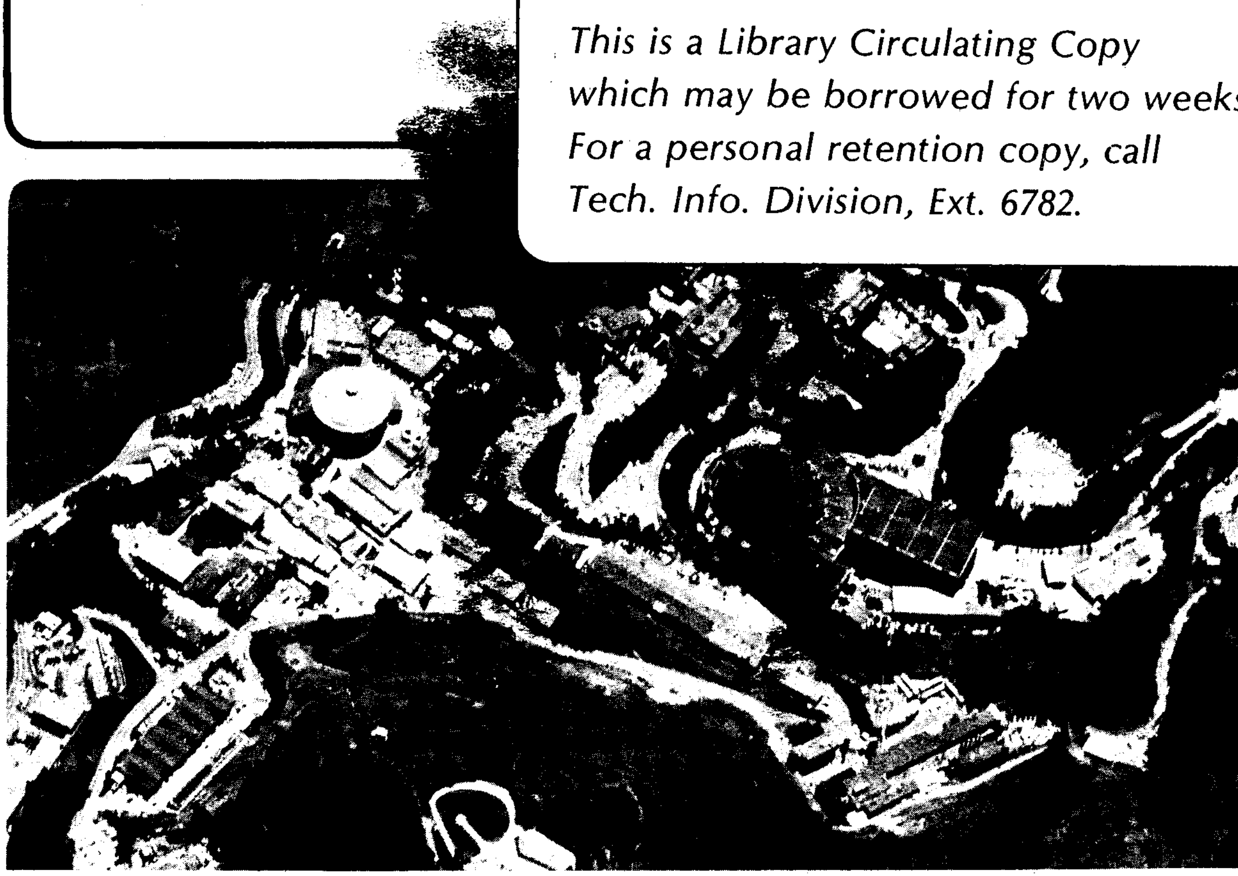
Lawrence Berkeley Laboratory
UNIVERSITY OF CALIFORNIA

Physics, Computer Science & Mathematics Division

REPEAT-FREE SEQUENCES

V. Elser

September 1983



Prepared for the U.S. Department of Energy under Contract DE-AC03-76SF00098

September 1983 LBL-16632

REPEAT-FREE SEQUENCES*

Veit Elser

Lawrence Berkeley Laboratory
and
Department of Physics
University of California
Berkeley, California 94720

ABSTRACT

A special class of sequences having no adjacent repeated subsequences is discussed. It is shown that "repeat-free" sequences of three symbols having arbitrary length exist. Moreover, the exponential growth of the number of such sequences with their length is established.

* This work was supported by the Director, Office of Energy Research, Office of High Energy and Nuclear Physics, Division of High Energy Physics of the U. S. Department of Energy under Contract DE-AC03-76SF00098.

A repeat-free sequence is a sequence of letters $A, B, C, \ldots$ that has no adjacent repeated subsequences. For example, $ABACAB$ is repeat-free whereas $AB\underline{ACAC}$ is not. Using only the three letters A, B and C, it will be shown there are repeat-free sequences of arbitrary length. This will involve a transformation that operates on repeat-free sequences by replacing each letter by a "word" and leaves the new sequence repeat-free. Repeated application of this transformation generates arbitrarily long repeat-free sequences. A particular example of such a transformation is:

$$A \rightarrow A' = ABCBACBCABCBA$$

$$B \rightarrow B' = BCACBACABCACB$$

$$C \rightarrow C' = CABACBABCABAC$$

Here the words are of length 13. In general, each of the letters will be replaced by a word of length $p > 1$. A sufficient condition on the words A', B', and C' that insures the new sequence of letters will be repeat-free is the following[1] :

(i) X' begins and ends with X for all $X \in \mathcal{S} = \{A, B, C\}$

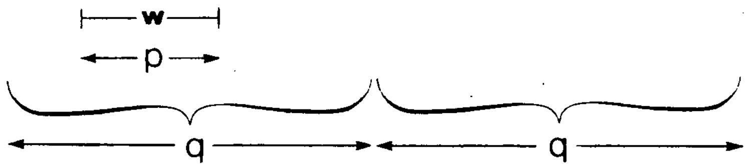
(ii) the sequence of length $3p$,

$$X'Y'Z'$$

is repeat-free for all $X \in \mathcal{S}, Y \in \mathcal{S}, Z \in \mathcal{S}$ with $X \neq Y, Y \neq Z$.

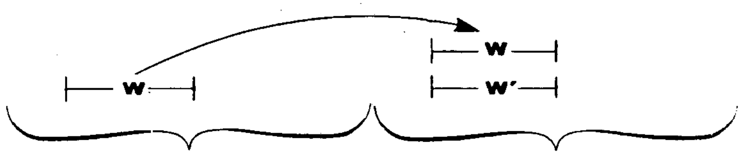
Proof: Suppose the sequence S is repeat-free. Then S' is obtained from S by a transformation as given above. S' is also a sequence of words of length p. We will see that the hypothesis that S' has a repeat leads to a contradiction.

First, suppose one of the repeated subsequences entirely contains one of the words of S'. Without loss of generality[2], assume it is the first(leftmost) subsequence:

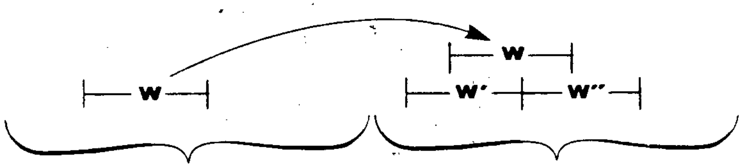


As shown, the repeat has length q. Also, let **w** be the first such word (no complete word contained in the repeat to its left). By assumption, this word perfectly matches a length p sequence of letters in the second subsequence. Two situations are possible[3]:

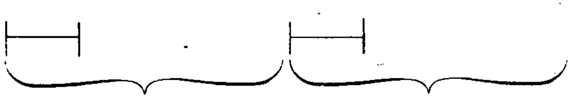
(1)



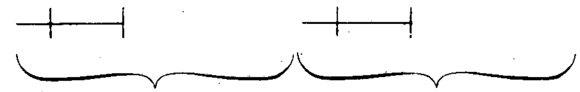
(2)



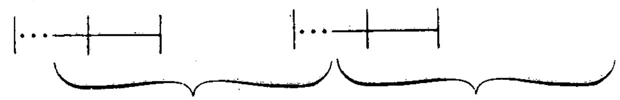
In (1) **w**=**w**′ and these either appear at the beginning of the repeated subsequences,

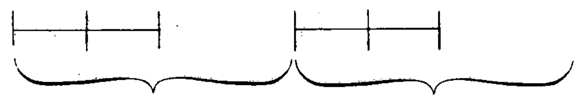


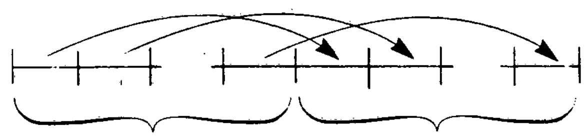
or there is a word fragment separating them from the beginning of the subsequences,



In the latter case, the matching of the word fragments implies the matching of their last letters which by (i) implies the fragments belong to the same word. Thus, the fragments can be extended into whole words yielding the former situation:

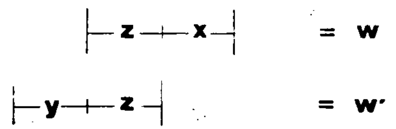




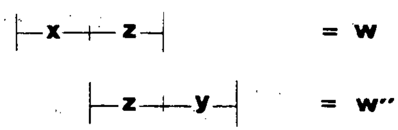


But this is clearly impossible since this implies the original sequence S had a repeat.

Now consider case (2). We cannot have $\mathbf{w}=\mathbf{w}'$ and $\mathbf{w}=\mathbf{w}''$ since then $\mathbf{w}'=\mathbf{w}''$ which implies that S was not repeat-free (it would have had two repeated letters). Suppose $\mathbf{w}\neq\mathbf{w}'$. Since fragments of these match we can write,

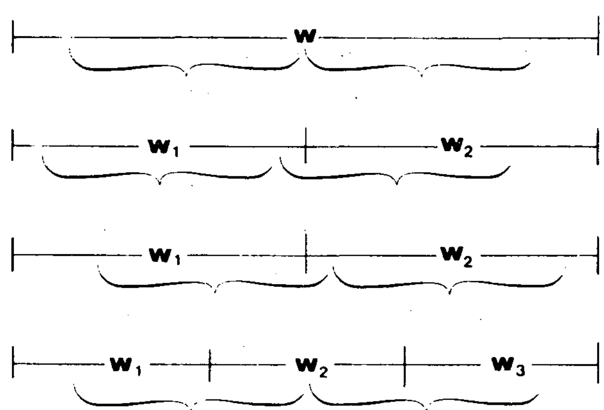


This contradicts (ii) since now $\mathbf{w}'\mathbf{w}=\mathbf{yzzx}$ has a repeat (as would $\mathbf{w}'\mathbf{wv}$ for any $\mathbf{v}\neq\mathbf{w}$). On the other hand, if $\mathbf{w}\neq\mathbf{w}''$ then,



and $\mathbf{ww}''=\mathbf{xzzy}$ has a repeat.

Finally, consider the situation where neither of the repeating subsequences entirely contains a word of S'. Now the following may arise:



Clearly, each subsequence consists of at most two word fragments. Also, if both contain two fragments then the subsequences contain fragments coming from a common word (the last case, above). Thus the two substrings together are made up of fragments coming from at most three consecutive words $\mathbf{w}_1\mathbf{w}_2\mathbf{w}_3$. Moreover, we know that $\mathbf{w}_1\neq\mathbf{w}_2$ and $\mathbf{w}_2\neq\mathbf{w}_3$ otherwise S had a repeat. But this contradicts (ii) since we are saying $\mathbf{w}_1\mathbf{w}_2\mathbf{w}_3$ has a repeat.QED

The transformation rule given earlier can be generalized in a way that permits the construction of exponentially many repeat-free sequences (the proof is trivially modified[4]) :

$$A \to A_{k_A} \in \{A_1, A_2, \ldots, A_{n_A}\}$$

$$B \to B_{k_B} \in \{B_1, B_2, \ldots, B_{n_B}\}$$

$$C \to C_{k_C} \in \{C_1, C_2, \ldots, C_{n_C}\}$$

where the words X_{k_X}, $X \in S = \{A, B, C\}$, $1 \leq k_X \leq n_X$ are all of length $p > 1$ and satisfy:

(i′) X_{k_X} begins and ends with X for all $X \in S$, $1 \leq k_X \leq n_X$

(ii′) The sequence of length $3p$

$$X_{k_X} Y_{k_Y} Z_{k_Z}$$

is repeat-free for all

$$X \in S, \quad Y \in S, \quad Z \in S$$

$$1 \leq k_X \leq n_X, \quad 1 \leq k_Y \leq n_Y, \quad 1 \leq k_Z \leq n_Z$$

with $X \neq Y, Y \neq Z$.

For the case $n_A = n_B = n_C = n$, a repeat-free sequence of length L may be transformed into any of n^L distinct repeat-free sequences of length nL. Also, it is easy to see that for different sequences of choices of the substitution words (ie., $(k_A, k_B, k_C)_1$, $(k_A, k_B, k_C)_2$, ...) a sequence is never generated twice (consider property (i′) which is independent of the choices k_X). Thus if $\#(L)$ counts the number of repeat-free sequences of length L that can be generated in this way, we have:

$$\#(1) = 3$$
$$\#(p) = 3n^1$$
$$\#(p^2) = 3n^1 n^p$$
$$\#(p^3) = 3n^1 n^p n^{p^2}$$
$$\vdots$$
$$\#(p^r) = 3n^{\frac{p^r-1}{p-1}}$$
$$\#(L) \sim \left(n^{\frac{1}{p-1}}\right)^L$$

An example with $n = 2, p = 29$ is[5]:

$$A_1 = ABCACBCAB\underline{AC}ABCACBA\underline{CA}BACBCACBA$$

$$A_2 = ABCACBCAB\underline{CB}ABCACBA\underline{BC}BACBCACBA$$

with the other transformations obtained from this by cyclic permutation: $A \to B \to C \to A$.

Note that A_1 and A_2 are palindromes (reflection symmetric). This property plus the cyclic relationship of the transformations makes it only necessary to check that,

$$A_k B_l A_m \qquad A_k B_l C_m \qquad A_k C_l A_m$$

are repeat-free for all choices of the subscripts (to verify (ii′)).

This example gives an exponential lower bound on the number, $N(L)$, of repeat-free sequences of length L. That is, for L sufficiently large,

$$N(L) > \#(L) \sim \left(2^{1/28}\right)^L \approx (1.025)^L$$

Compared to this rather weak lower bound, it is very easy to obtain a descending sequence of upper bounds. Suppose we try to generate repeat-free sequences by writing down all possible sequences of repeat-free words of length p. This ignores possible repeats occuring among adjacent words, but avoids repeats within words. When a word is added to the sequence it must be compatible with the last two letters of the sequence which, without loss of generality, we take to be AB. Hence, we must count the number of repeat-free words, $\omega(p)$, of length $p + 2$ with initial letters AB. Now, since there are at most $\omega(p)$ choices whenever the sequence is lengthened by p, an upper bound on the growth rate of repeat-free sequences is $\omega(p)^{1/p}$. For example, taking $p = 5$ we have:

$(AB)ACABC$	$(AB)CACBA$
$(AB)ACABA$	$(AB)CACBC$
$(AB)ACBAB$	$(AB)CBABC$
$(AB)ACBCA$	$(AB)CBACA$
$(AB)CABAC$	$(AB)CBACB$

$$\omega(5) = 10 \qquad \omega(5)^{1/5} \approx 1.58$$

From the existence of bounds,

$$c_1^L < N(L) < c_2^L$$

on the number of repeat-free sequences of length L, we deduce that,

$$\lim_{L \to \infty} \mathcal{N}(L)^{1/L} = c$$

exists and,

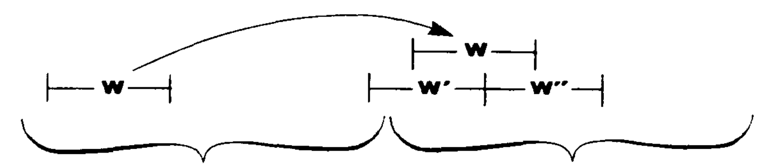
$$1.025 < c < 1.58$$

Footnotes

[1] I am indebted to Alan Goldman for pointing out to me a flaw in an earlier version of the proof based on a weaker form of condition (ii).

[2] Interchanging left and right gives the analysis of the other situation.

[3] In case (2) the word $\mathbf{w}'$ is drawn so it lies entirely within the second subsequence. It may also happen that,



However, the analysis of case (2) does not depend on this detail. All that matters is that $\mathbf{w}$ has a non-empty intersection with $\mathbf{w}'$and $\mathbf{w}''$.

[4] The only part that requires some attention is the step on page 3 where the initial word fragments are extended into full words. The identity of the last letters of these fragments no longer implies the identity of the actual words (they might be A_i and A_j with $i \neq j$). However, it still follows that the predecessors of these words are identical letters, leading to the same contradiction that S had a repeat.

[5] Note that the two sequences differ in only four places (underlined). In spite of their length, these were quite easy to find. I generated about 5 palindromic candidates by hand and tested them for property (ii$'$) by computer. The pair given was the first one found; not necessarily the shortest. It would be interesting to know the minimum values of p for $n = 1, 2, 3, \ldots$.

Acknowledgments

The author thanks Mark Zimmermann for communicating the problem and Alan Goldman for pointing out an error. This work was supported by the Director, Office of Energy Research, Office of High Energy and Nuclear Physics, Division of High Energy Physics of the U.S. Department of Energy under Contract DE-AC03-76SF00098.

TECHNICAL INFORMATION DEPARTMENT
LAWRENCE BERKELEY LABORATORY
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720