Universal sequences

Citation for published version (APA):
Hollmann, H. D. L., & van Lint, J. H. (1997). Universal sequences. *Applicable Algebra in Engineering, Communication and Computing*, *8*(5), 347-352. https://doi.org/10.1007/s002000050071

DOI:
10.1007/s002000050071

Document status and date:
Published: 01/01/1997

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

AAECC **8**, 347–352 (1997)

AAECC
Applicable Algebra in
Engineering, Communication
and Computing
© Springer-Verlag 1997

Universal Sequences

Henk D. L. Hollmann[1], **J. H. van Lint**[1,2]

[1] Philips Research Laboratories, Prof. Holstlaan 4, NL-5656 AA Eindhoven, The Netherlands
(e-mail: hollmann@natlab.research.philips.com)
[2] Eindhoven University of Technology, Eindhoven, The Netherlands
(e-mail: wsdwjhvl@urc.tue.nl)

Dedicated to Amio Tietäväinen on the occasion of his 60th birthday

Received: November 4, 1996

Abstract. An (n, k)-universal sequence is a binary sequence with the property that each window of size k and span at most n is covered by the sequence, i.e., each sequence of length k occurs as the content of a shift of the window. We derive upper and lower bounds on the minimum length of universal sequences, both for the linear case and the circular case.

Keywords: Universal sequence, De Bruijn sequence, Paley sequence.

1 Introduction

In this paper we consider (0, 1)-sequences $X = x_0, x_1, \ldots, x_{L-1}$. We call L the *length* of the sequence. An increasing sequence of indices $i_1, i_2, \ldots, i_k$ with $i_k = i_1 + m - 1$ is called a *window* of *span* m and *size* k. We also use the name (m, k)-window. The index i_1 is called the initial position of the window. The subsequence $x_{i_1}, x_{i_2}, \ldots, x_{i_k}$ of X is called the *contents* of the window.

We call the sequence X an (n, k)*-universal sequence* if for every m with $k \leqq m \leqq n$ and for every window $0 = w_1, w_2, \ldots, w_k = m - 1$, each vector $\boldsymbol{a} \in \{0, 1\}^k$ occurs somewhere as the contents of the *shifted* window $w_1 + j$, $w_2 + j, \ldots, w_k + j$. (Here $j \leqq L - 1 - w_k$.) This terminology (with a slightly different meaning) was introduced by A. Lempel and M. Cohn in [2]. Such sequences have applications in the testing of very large scale integration (VLSI) chips.

Universal sequences are in some sense a generalization of the well known De Bruijn sequences. Recall that a De Bruijn sequence of length $L = 2^k$ is an arrangement of a (0,1)-sequence $x_0, x_1, \ldots, x_{L-1}$ on a circle such that the 2^k windows $i, i + 1, \ldots, i + k - 1$ (where we use the convention $x_i = x_{i+L}$) contain all possible vectors in $\{0, 1\}^k$. An example is $X = 0, 0, 0, 1, 1, 1, 0, 1$. From this we see

Correspondence to: J. H. van Lint (first address) that $0,0,0,1,1,1,0,1,0,0$ is a (3,3)-universal sequence of length 10. This is optimal since we obviously have $L \geqq n + 2^k - 1$ for an (n,k)-universal sequence of length L because we need at least 2^k initial positions for every (m,k)-window.

We are interested in the minimal length of an (n,k)-universal sequence. We denote this length by $L(n,k)$ and we define $f_k(n)$ by $L(n,k) = n + f_k(n)$. As observed above

$$f_k(n) \geqq 2^k - 1. \tag{1}$$

From the existence of De Bruijn sequences we find that for $n = k$ equality holds in (1).

We shall also study the circular generalization. Again, the sequence $x_0, x_1, \ldots, x_{L-1}$ is placed on a circle and indices (in subsequences and in windows) are considered $\bmod L$. We define $L^*(n,k)$ to be the minimal length of a circular universal (n,k)-sequence. So we have $L^*(k,k) = 2^k$.

The following restriction is also of interest. We define $M(n,k)$ to be the minimal length of a (0, 1) sequence that has the universality property for all (m,k)-windows with $m = n$.

In our analysis of (n,k)-universal sequences of length L, we shall often use the following array:

$$\chi := \begin{pmatrix} x_0 & x_1 & \cdots & x_{n-1} \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & & \vdots \\ x_{L-n} & x_{L-n+1} & \cdots & x_{L-1} \end{pmatrix}. \tag{2}$$

We call this the *array* χ of X. The columns of χ are called $\boldsymbol{x}^0$ to $\boldsymbol{x}^{n-1}$. The universality property implies that if we take k columns $\boldsymbol{x}^0 = \boldsymbol{x}^{i_1}, \boldsymbol{x}^{i_2}, \ldots, \boldsymbol{x}^{i_k} = \boldsymbol{x}^{n-1}$, then the submatrix of χ consisting of these columns contains all possible vectors in $\{0,1\}^k$ as rows. In fact, this is true for the restricted case with windows of span n only (i.e., when we consider $M(n,k)$).

As a first example of the use of this matrix we prove a lower bound for the length of universal sequences.

Theorem 1 *For* $k \geqq 4$ *we have* $f_k(n) \geqq \log_2 n - 1$.

Proof. By the argument above, the four columns $\boldsymbol{x}^0, \boldsymbol{x}^i, \boldsymbol{x}^j$ and $\boldsymbol{x}^{n-1}$ (where $0 < i < j < n-1$) must be different. It follows that $n - 2 \leqq 2^{L-n+1} - 2$ and this proves the assertion. □

Remark. Note that we have in fact shown that

$$M(n,k) \geqq n + \log_2 n - 1.$$

Remark. This method does not work for $k = 3$. In fact, we can show by a direct construction that $M(n,3) \leqq n + 15$ for $n \geqq 28$ (we omit this here). It is unlikely that such result holds for $L(n,3)$ but we have not been able to show that $f_3(n)$ is not bounded. The difficulty for the case $k = 3$ is demonstrated by the following argument. Suppose 'that $L(n,3) \leqq n + c$ for some constant c and suppose that there is a sequence $x_0, x_1, \ldots$ such that for each n, the initial part $x_0, x_1, \ldots, x_{n-1+c}$ is universal. Consider the corresponding array χ. There are only 2^c possible columns for χ. Then for every integer M there are indices i and j in the interval $[M, M + 2^c]$ such that the columns $\boldsymbol{x}^i$ and $\boldsymbol{x}^j$ are equal. This implies that the first c shifts of the window $(0, i, j)$ do not cover all possible triples. But then the initial part $x_0, x_1, \ldots, x_{j-1+c}$ is not universal, a contradiction.

2 $(n, 2)$-universal sequences

The case $k = 2$ is almost trivial. We already know that $f_2(2) = 3$. If a $(3, 2)$-universal sequence of length 6 exists, we see from its array that it must have two 0's and two 1's in the first four positions and also in the last four positions. Furthermore, it must contain two consecutive 0's and two consecutive 1's. Only six $(0, 1)$ sequences satisfy these conditions and none of them is universal. So, $L(3, 2) \geqq 7$.

Theorem 2 *We have* $L(n, 2) = n + 3$ *for* $n \geqq 4$.

Proof. Consider the sequence X starting with $0, 0, 1$ and continuing with 1 and 0 alternating. Its array χ has $x^0 = (0, 0, 1, 1)^\top$ and all columns x^j with $j \geqq 3$ have 0 and 1 alternating. Hence the four pairs (x_i, x_{i+j}), $0 \leqq i \leqq 3$ are different. This handles all $(m, 2)$-windows of span > 3. By inspection all $(m, 2)$-windows of span $m = 2$ and $m = 3$ also contain every possible vector of length 2. So we are done and we have also shown that $L(3, 2) = 7$. □

Corollary $L^*(n, 2) = n$ *for* $n \geqq 7$.

Proof. We use the same sequence as above (now with length $n \geqq 7$). The array χ is the same as before with the exception of the last three columns (because the sequence is circular). So now all $(m, 2)$-windows with $m \leqq n - 3$ have the required property. However, by reversing the order of the two elements in a window, this implies that all windows with $m \geqq 5$ also have the required property. □

Remark. From the De Bruijn sequence 0, 0, 1, 1 we have $L^*(2, 2) = 4$. For a circular universal $(3, 2)$-sequence we must have two adjacent 0's and two nonadjacent 0's and similarly for 1's. So the length must be at least 6. Then the sequence 0, 0, 1, 1, 0, 1 shows that $L^*(n, 2) = 6$ for $3 \leqq n \leqq 6$.

3 An upper bound

We shall now show that there is a constant c_k such that $f_k(n) \leqq c_k \log n$. Let $g_k(n) := 2^{k-1} k^3 \log(2n)$. We shall show that $f_k(n) \leqq g_k(n)$. The idea is to show that for each window it is possible to find sufficiently many shifted versions that are pairwise disjoint. Subsequently, we consider all possible $(0, 1)$ sequences of the required length and delete those that do not cover all $(0, 1)$ vectors in the shifted windows. By showing that not all the sequences are deleted in this way, we establish the existence of a universal sequence.

Consider a *fixed* window $W := \{0 = w_0, w_1, \ldots, w_{k-1} = m - 1\}$ of span m and size k. The shifted windows $a_i + W$, where $0 = a_0 < a_1 < \ldots < a_{r-1}$ are disjoint if for all i and j $(i \neq j)$ the difference $a_i - a_j$ is not equal to some difference $w_\mu - w_\nu$ with w_μ and w_ν in W. Now $w_\mu - w_\nu$ takes on at most $\frac{1}{2}k(k - 1)$ positive values. So, if $a_0, a_1, \ldots, a_{i-1}$ satisfy the constraints, there are at most $i(1 + \frac{1}{2}k(k - 1))$ excluded values for a_i. Therefore, a sequence $a_0, a_1, \ldots, a_{r-1}$ such that the shifted windows are disjoint can be found, with

$$a_{r-1} \leqq (r - 1)(1 + \tfrac{1}{2}k(k - 1)). \tag{3}$$

We shall use the following trivial lemma.

Lemma 1 *If A is an alphabet of size a, then among all sequences $(\xi_1, \xi_2, \ldots, \xi_r) \subset A^r$ there are at most $a(a-1)^r$ sequences in which some element of A does not occur.*

Consider the set $\mathscr{L}$ of all sequences $x_0, x_1, \ldots, x_{L-1}$ in $\{0,1\}^L$ where $L := n + g_k(n)$. The r shifts $W + a_0, W + a_1, \ldots, W + a_{r-1}$ of the window W are pairwise disjoint k-tuples. Here by (3) the index r satisfies

$$r \geqq \frac{2g_k(n)}{k^2}. \tag{4}$$

By Lemma 1 there are at most

$$2^{L-kr} \cdot 2^k (2^k - 1)^r \tag{5}$$

sequences $x_0, x_1, \ldots, x_{L-1}$ such that some vector $\mathbf{c} \in \{0,1\}^k$ is missing among the contents of the r shifted windows.

We delete these (0, 1) sequences from $\mathscr{L}$ and in fact to this for *every* window W of span $\leqq n$ and size k. The number of such windows is $\binom{n-1}{k-1} < n^k$. We see from (5) that after all the deletions there remains a universal sequence if

$$2^L > n^k \cdot 2^{L-kr} \cdot 2^k (2^k - 1)^r,$$

i.e. if

$$1 > (2n)^k \left(1 - \frac{1}{2^k}\right)^r. \tag{6}$$

By (6) we are done if $r > 2^k \cdot k \log(2n)$ and by (4) this is true. This completes the proof of the following theorem.

Theorem 3 *For every $k \geqq 3$ there is a constant c_k* such that

$$L(n,k) \leqq n + c_k \log n.$$

Remark. After completion of this work, we became aware of [4]. Here, the authors investigate (n,k)-universal *test sets*, $N \times n$ matrices T with the property that on any k-tuple of coordinates each of the 2^k possible vectors occurs at least once. The number of rows N is called the size of the test set. Moreover, they call a sequence $X = x_0, \ldots, x_{L-1}$ (n,k)-universal if the array χ of X is an (n,k)-universal test set. (So their definition is slightly stronger than ours.) In that paper Theorem 3 is also obtained, with a similar proof.

4 The circular case

We first consider some small cases of $L^*(n,3)$. Clearly $L^*(3,3) = 8$ and in fact the sequence is unique, namely 0, 0, 0, 1, 1, 1, 0, 1. This sequence does not contain a window $(i, i+1, i+3)$ with contents 0, 0, 0. Therefore $L^*(4,3) \geqq 9$. A universal sequence must contain three adjacent 0's and three adjacent 1's. Assume that $L^*(4,3) = 9$. We distinguish two cases:

(i) There are four adjacent 0's. This is possible in only one way, namely 0, 0, 0, 0, 1, 1, 1, 0, 1 and this sequence does not contain a window $(i, i+2, i+3)$ with contents 0, 1, 0.

(ii) No four adjacent 0's or 1's. Without loss of generality we now can assume that the sequence is 0, 0, 0, 1, 1, 1, 0, 1, 1. Now we do not have 0, 1, 0 as a consecutive subsequence.

This argument shows that $L^*(4,3) \geqq 10$. Then the sequence

$$1,0,1,1,1,0,1,0,0,0$$

shows that $L^*(4,3) = L^*(5,3) = 10$.

Arguments like this become increasingly difficult. We calculated some values of $L^*(n,3)$ for small n by computer. We found $L^*(6,3) = 12$, $L^*(7,3) = 14$, $L^*(8,3) = 16$, $L^*(9,3) = 17$, $L^*(n,3) = 18$ for $10 \leqq n \leqq 12$, and $L^*(n,3) = 19$ for $13 \leqq n \leqq 19$. Note that on a circular sequence of length 19 each window of size 3 can be viewed as one with length at most 13 (by changing the initial position).

The value for $n = 19$ is achieved by a Paley sequence: $x_i = 0$ if i is a square in $\boldsymbol{F}_{19}$ and $x_i = 1$ otherwise. This leads to our next theorem. We aim to show that for a fixed k there is a bound p_k such that for all primes $p \geqq p_k$ we have $L^*(p,k) = p$. For $k = 2$ this is easy. Let χ be the quadratic character on $\boldsymbol{F}_p$. We now use $\{+1, -1\}$ as alphabet. We define χ' by $\chi'(0) = 1$, $\chi'(a) = \chi(a)$ for $a \neq 0$. We claim that the sequence $x_i := \chi'(i)$, $0 \leqq i < p$, is a circular universal sequence for $k = 2$ if p is sufficiently large. To show this, we use the following well known fact (cf. [3], Ch. 18). For any $c \neq 0$ in $\boldsymbol{F}_p$ we have

$$\sum_{b \in \boldsymbol{F}_p} \chi(b)\chi(b+c) = -1. \tag{7}$$

Since χ takes on the values $+1$ and -1 exactly $\frac{1}{2}(p-1)$ times it easily follows from (7) that the pair $(\chi'(b), \chi'(b+c))$ takes on each of the four values $(+1, +1)$, $(+1, -1)$, $(-1, +1)$, and $(-1, -1)$ roughly $\frac{1}{4}p$ times. In fact, for each pair the deviation from $\frac{1}{4}p$ is at most 2. This proves the universality (for $p \geqq 11$; in fact, for $p = 5$ and $p = 7$ it is also true).

We shall proceed by induction. We need a lemma to estimate sums similar to the one in (7) (see e.g. [1, Theorem 5.41]).

Lemma 2 *Let ψ be a multiplicative character of $\boldsymbol{F}_q$ of order $m > 1$ and let $f \in \boldsymbol{F}_q[x]$ be a monic polynomial of positive degree that is not an m-th power of a polynomial. Let d be the number of distinct roots of f in its splitting field over $\boldsymbol{F}_q$. Then for every $a \in \boldsymbol{F}_q$ we have*

$$\left| \sum_{c \in \boldsymbol{F}_q} \psi(af(c)) \right| \leqq (d-1)q^{\frac{1}{2}}.$$

We shall show that for a long Paley sequence the circular shifts of a window of size k contain every possible sequence roughly $p/2^k$ times. This is formulated as a lemma.

Lemma 3 *For any k there are constants c_k and d_k such that for all primes $p > k$ the following holds. For any (m,k)-window $w_1, w_2, \ldots, w_k$ of span $\leqq p$, the circular shifts of this window along the sequence $\chi'(i)$ have every possible vector in $\{0,1\}^k$ as contents $p/2^k + \varepsilon$ times, where for each of the possible contents the deviation ε satisfies*

$$|\varepsilon| \leqq c_k\sqrt{p} + d_k. \tag{8}$$

We have shown that Lemma 3 is true for $k = 2$. We apply Lemma 2 with $q = p$ and $\psi = \chi$ to the function

$$f(z) := (z + w_1)(z + w_2) \cdots (z + w_k).$$

Then $d = k$. Take $a = 1$. We find

$$\left| \sum_{c \in \boldsymbol{F}_q} \chi(c + w_1)\chi(c + w_2) \cdots \chi(c + w_k) \right| \leqq (k - 1)\sqrt{p}. \tag{9}$$

If we replace χ by χ', the right hand side of (9) increases by at most k. For any $e = (e_1, e_2, \ldots, e_k) \in \{+1, -1\}^k$ let $n_{\boldsymbol{e}}$ denote the number of occurrences of $\boldsymbol{e}$ as the contents of a shifted window. Then we can read (9) as

$$\left| \sum_{\boldsymbol{e} \in \{+1, -1\}^k} ((e_1 e_2 \ldots e_k) n_{\boldsymbol{e}} \right| \leqq (k - 1)\sqrt{p} + k. \tag{10}$$

If $\boldsymbol{e}$ and $\boldsymbol{f}$ are two vectors in $\{+1, -1\}^k$ that differ in only one coordinate, then the induction hypothesis states that

$$n_{\boldsymbol{e}} + n_{\boldsymbol{f}} = \frac{p}{2^{k-1}} + r, \tag{11}$$

where the remainder term r depends on the pair but has an absolute value at most $c_{k-1}\sqrt{p} + d_{k-1}$ with certain constants c_{k-1} and d_{k-1}. Each term $n_{\boldsymbol{e}}$ can be written as a linear combination of the left hand side of (10) and a number of terms of the type occurring in the left hand side of (11). We omit the details of this elementary linear algebra which produces the assertion of Lemma 3 by induction.

From Lemma 3 we see that if p is sufficiently large, all vectors indeed occur at least once as contents of a shifted window.

Theorem 4. *For any k there is a p_k such that for all primes $p > p_k$ the sequence X defined by $x_i := \chi'(i)$ for $0 \leqq i < p$ is a circular universal sequence.*

This shows that for fixed k the function $L^*(n, k)$ is asymptotically equal to $n (n \to \infty)$. Computer results suggest existence of a number n_k such that $L^*(n, k) = n$ for $n \geqq n_k$. We have shown that $n_1 = 2$ and $n_2 = 5$. Probably $n_3 = 19$, $n_4 = 67$ and $n_5 = 331$, but we have not proved this.

References

1. Lidl, R., Niederreiter, H.: Finite fields, Reading MA: Addison-Wesley 1983
2. Lempel, A., Cohn, M.: Design of universal test-sequences for VLSI. IEEE Trans. Inform. Theory **IT-31**, 10–17 (1985)
3. van Lint, J. H., Wilson, R. M.: A course in combinatorics. Cambridge: Cambridge University Press, 1992
4. Seroussl, G., Bshouty, N. H.: Vector sets for exhaustive testing of logic circuits. IEEE Trans. Inform. Theory **IT-34**, 513–522 (1988)