SIAM J. DISC. MATH.
Vol. 7, No. 4, pp. 598–604, November 1994

008

ON UNIVERSAL CYCLES FOR k-SUBSETS OF AN n-SET*

GLENN HURLBERT†

Abstract. A *universal cycle,* or *Ucycle,* for k-subsets of $[n] = \{1, \dots, n\}$ is a cyclic sequence of $\binom{n}{k}$ integers with the property that each subset of $[n]$ of size k appears exactly once consecutively in the sequence. Chung, Diaconis, and Graham have conjectured their existence for fixed k and large n when $n | \binom{n}{k}$. Here the Ucycles for $k = 3, 4, 6$ and large n relatively prime to k are exhibited.

Key words. universal cycle, de Bruijn cycle

AMS subject classifications. 05B30, 05C38, 05C45, 94A55

1. Introduction. In 1894 the following binary sequence was discovered by Flye-Sainte Marie [M]:

$$00011101.$$

The interesting property of this sequence is that, when regarded as a cyclic sequence, every binary triple appears uniquely as a consecutive block of digits. Indeed, 000, 001, 011, 111, 110, 101, 010, 100 is the corresponding list of binary triples. It was also discovered that for all n and k analogous k-ary sequences exist listing k-ary n-tuples uniquely. This result went largely unnoticed until the sequences were rediscovered in 1946 by de Bruijn [B] and Good [G], and they have come to be known as *de Bruijn cycles.* For interesting results, generalizations, applications, and history see [F], [H].

Recently, Chung, Diaconis, and Graham [C] have generalized such sequences so as to list combinatorial families other than k-ary n-tuples including permutations of $[n]$, partitions of $[n]$, k-subsets of $[n]$, and k-dimensional subspaces of an n-dimensional vector space over a finite field. They call such sequences *universal cycles,* or *Ucycles.*

In this paper we discuss only the case of Ucycles for k-subsets of $[n]$. An example with $k = 3, n = 8$ is as follows.

$$1356725 \quad 6823472 \quad 3578147 \quad 8245614 \quad 5712361 \quad 2467836 \quad 7134583 \quad 4681258$$

The list of subsets begins with $135, 356, 567, \dots,$ and ends with $\dots, 258, 581, 813$. (Often, we will leave out the brackets in set notation to improve readability.) And, of course, there are $\binom{8}{3} = 56$ digits in the sequence, one for each subset. In the example above, spaces were used only to help the reader notice that each block is just an additive translation of the previous block. That is, we add 5 modulo 8 to the digits of one block to obtain the digits of the next block. This will be an important feature of Ucycle construction.

We first remark that any ordering of $[n]$ is a Ucycle both for 1- and $(n-1)$-subsets. These we call trivial Ucycles. Also, any listing of the vertices of an Eulerian circuit in the complete graph K_n (n *odd*) is a Ucycle for 2-subsets of $[n]$. Thus, we shall always assume that $k \geq 3$ and $n \geq k + 2$.

Next, we notice that each integer must occur equally often in a Ucycle since each integer is in equally many k-subsets of $[n]$. So for the existence of such Ucycles we

* Received by the editors August 6, 1991; accepted for publication November 8, 1993.

† Department of Mathematics, Arizona State University, Tempe, Arizona 85287-1804.

have the following necessary condition

$$\text{(NC)} \qquad n \text{ divides } \binom{n}{k}.$$

Equivalently, we may write that k divides $\binom{n-1}{k-1}$. This carries the interpretation that any digit in the Ucycle is in k k-subsets and any integer is in $\binom{n-1}{k-1}$ k-subsets, so $\binom{n-1}{k-1}/k$ is the number of occurrences of each integer in the Ucycle. Notice that (NC) holds whenever n is relatively prime to k. Indeed, $\binom{n}{k} = n\binom{n-1}{k-1}/k$ is an integer and n and k have no common factors, so $\binom{n-1}{k-1}/k$ must be an integer. Also notice that (NC) fails whenever n is a multiple of k.

2. Results. In [C] Chung, Diaconis, and Graham made the following conjecture, for which $100 is offered.

CONJECTURE 1. *For all k there is an integer $n_0(k)$ such that, for $n \geq n_0(k)$, Ucycles for k-subsets of $[n]$ exist if and only if* (NC) *holds.*

In [J1] we find the following results.

THEOREM 2. *Ucycles for* 3*-subsets of $[n]$ exist for all $n \geq 8$ not divisible by* 3.

THEOREM 3. *Ucycles for* 4*-subsets of $[n]$ exist for all odd $n \geq 9$.*

Theorem 2 is a verification of conjecture 1 for $k = 3$.

In order to verify the conjecture for $k = 4$ one would need to construct Ucycles for $n \equiv 2 \bmod 8$. No examples of this case have ever been found, nor have any with $k = 5$. The purpose of this paper is to prove the following theorem.

THEOREM 4. *For $k = 3, 4, 6$ there is an integer $n_0(k)$ such that, for $n \geq n_0(k)$, Ucycles for k-subsets of $[n]$ exist whenever n is relatively prime to k.*

This theorem includes both results of Jackson and unifies them in one setting together with the case $k = 6$.

3. Terminology. Let $S = \{s_1, \ldots, s_k\}, s_i < s_{i+1}$, be a k-subset of $[n]$. Define its *ordered difference set*, or d-set, $D(s) = (d_1, \ldots, d_k)$ by $d_i = s_{i+1} - s_i$, where indices are modulo k and arithmetic is modulo n. Two d-sets are *equivalent* if one is a cyclic permutation of the other. Thus, we say that any additive translation of S belongs to the same d-set as S, where an *additive translation* of S is any set $S + r = \{s_1 + r, \ldots, s_k + r\}$ modulo n. Two d-sets will belong to the same *d-class* whenever one is any permutation of the other. Thus the family of all d-classes is simply the family of all unordered partitions of the integer n into k parts. Each d-class in turn defines a partition of k according to the number of parts of the same size. We say that two d-classes belong to the same *d-pattern* if they define the same partition of k. A d-pattern is *good* if some part has size 1 and *bad* otherwise. Let us offer some examples with $k = 5$ and $n = 40$.

There are 7 d-patterns $\langle 1,1,1,1,1\rangle, \langle 2,1,1,1\rangle, \langle 3,1,1\rangle, \langle 4,1\rangle, \langle 5\rangle, \langle 2,2,1\rangle$, and $\langle 3,2\rangle$, of which only $\langle 5\rangle$ and $\langle 3,2\rangle$ are bad. $\langle 3,2\rangle$ contains the d-classes $[2,2,2,17,17], [4,4,4,14,14], \ldots$, and $[12,12,12,2,2]$. $[2,2,2,17,17]$ contains the d-sets $(2,2,2,17,17)$ and $(2,2,17,2,17)$. $(2,2,17,2,17)$ contains the sets $\{1,3,5,22,24\}$, $\{2,4,6,23,25\}, \ldots$, and $\{40,2,4,21,23\}$. The notation of braces, parentheses, brackets, and angles will be maintained throughout to distinguish the objects from one another.

4. Proof of Theorem 4. We first prove the following lemma.

LEMMA 5. *No bad d-patterns exist if and only if $k = 3, 4$, or* 6 *and* $\gcd(n, k) = 1$.

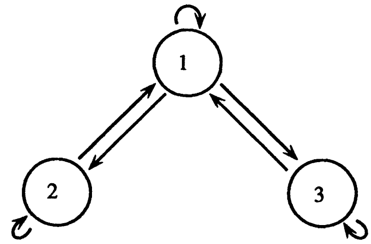

FIG. 1. $\mathcal{T}_{8,3}$.

Proof. If $r = \gcd(n,k)$ then the d-pattern $\langle r, r, \ldots, r\rangle$ is bad unless $r = 1$. If $k > 3$ is odd then $\langle t, t+1\rangle$ is bad, where $k = 2t+1$ and $t > 1$, and if $k > 6$ is even then $\langle 3, t, t+1\rangle$ is bad, where $k = 2t+4$ and $t > 1$. Finally, it is easy to verify that if $k = 3, 4$, or 6 and $\gcd(n,k) = 1$ then no bad d-patterns exist. □

Next we define the *transition graph* $\mathcal{T}_{n,k}$, which is dependent upon the choices of representations for each d-class. Given the d-set $(d_1, \ldots, d_{k-1}, d_k)$ we distinguish one of its coordinates (which we may assume to be d_k because of cyclic permutations) in the representation $(d_1, \ldots, d_{k-1}; d_k)$ so as to imply the ordering $\{i, i+d_1, \ldots, i+d_1+\cdots+d_{k-1}\}$ of all its sets. The use of the semicolon in the representation is to identify which of the k differences is not to be used.

The representation of a d-class by $[d_1, \ldots, d_{k-1}; d_k]$ signifies that d_k is distinguished (unused) in each of its d-sets. Thus, to avoid ambiguity, it is important that d_k be unique, in particular that the corresponding d-pattern be good. For example, with $k = 4$ and $n = 11$, we can choose $[2,2,1;6]$ to represent $[2,2,1,6]$. This determines the representations $(2,2,1;6), (2,1,2;6)$, and $(1,2,2;6)$ of its three d-sets, of which $(2,1,2;6)$ denotes the (ordered) sets $\{1,3,4,6\}, \{2,4,5,7\}, \ldots$ and $\{11,2,3,5\}$.

Given the d-set representation $(d_1, \ldots, d_{k-1}; d_k)$ we call the terms $((d_1, \ldots, d_{k-2}))$ its *prefix* and $((d_2, \ldots, d_{k-1}))$ its *suffix*. We use double parentheses to denote that these are vertices in the transition graph $\mathcal{T}_{n,k}$ whose directed edges are precisely the representations involved.

For example, Fig. 1 shows the transition graph $\mathcal{T}_{8,3}$ which was used to construct the Ucycle for 3-subsets of [8] above. The d-sets are represented by $(1,1;6), (2,2;4), (3,3;2), (1,2;5), (2,1;5), (1,3;4)$, and $(3,1;4)$. The d-set $(2,1;5)$ corresponds to the directed edge $((2)) \to ((1))$, and so on. The Eulerian circuit 2211331 corresponds to a listing of all d-sets and produces the differences in the first block, 1356725, along with the first digit, 6, of the next block. Since the sum $2+2+1+1+3+3+1 \equiv 5 \bmod 8$, each block shifts by 5, and since 5 is relatively prime to 8 each integer occurs as the starting digit of some block. Hence, each 3-subset of [8] occurs exactly once. As we shall see, it is an unnecessary luxury that the sum (5) is relatively prime to n (8).

LEMMA 6. *If $\mathcal{T}_{n,k}$ is Eulerian for some choice of representations of d-classes, then there exists a Ucycle for k-subsets of* $[n]$.

Proof of Lemma 6. $((1,1,\ldots,1))$ is always a vertex of $\mathcal{T}_{n,k}$ and there is always a loop at that vertex representing the d-set $(1,1,\ldots,1;d)$. This is because there is no other way to represent the d-class $[1,1,\ldots,1,d]$, because we would be unable to distinguish which 1 would be set apart by a semicolon if we tried. Let $S_1 = \{1,2,\ldots,k\}, S_2 = \{2,3,\ldots,k+1\} = S_1+1$, and in general, let $S_{i+1} = S_i + 1$ be the sets belonging to this representation. Starting at S_i we follow along the edges of an Eulerian circuit in $\mathcal{T}_{n,k}$ until we return to our original loop. Repeating the loop now determines the set $S_i + r$ for some $0 \le r < n$. Retrace the circuit repeatedly until we finally reach S_i again, now having used the sets $S_i, S_i+r, S_i+2r, \ldots,$ and $S_i - r$, in that order. Call the cycle produced U_i. Now let $s = \gcd(r, n)$. Then if $s = 1$ we have actually produced our Ucycle since we have used each of the n sets from every d-set (our initial example with $n = 8$ and $k = 3$ has $r = 5$ and $s = 1$). Otherwise, we have constructed s disjoint cycles $U_1, U_2, \ldots, U_s$, which must somehow be hooked together to form one. We do this by a method we call *insertion,* showing how to insert U_{i+1} into U_i. Actually, the general case is the same as the first, so we merely show how to insert U_2 into U_1. Given

$$U_1 = 1, 2, \ldots, k-1, k, x, \ldots, y$$

and

$$U_2 = 2, 3, \ldots, k, k+1, x+1, \ldots, y+1$$

we insert as below.

$$1, \underline{2, 3, \ldots, k, k+1, x+1, \ldots, y+1}, 2, 3, \ldots, k-1, k, x, \ldots, y$$

Induction then completes the proof. □

Next we take a closer look at $\mathcal{T}_{n,k}$ by defining the graph $\mathcal{T}_{n,k}(C)$ for each d-class $C = [d_1, \ldots, d_{k-1}; d_k]$. It is merely the restriction of $\mathcal{T}_{n,k}$ to edges that are representations of d-sets belonging to C. We now introduce the *class graph* $\mathcal{H}_{n,k}$, whose vertices are all possible d-classes and whose undirected edges join d-classes whose representations differ by one entry. For example, $[2, 2, 1, 3, 6; 7]$ and $[1, 1, 2, 2, 3; 12]$ are connected in $\mathcal{H}_{21,6}$.

LEMMA 7. *If $\mathcal{H}_{n,k}$ is connected and there are no bad d-patterns for k-subsets of $[n]$, then $\mathcal{T}_{n,k}$ is Eulerian.*

Proof of Lemma 7. If the d-class $C = [d_1, \ldots, d_{k-1}; d_k]$ belongs to a good d-pattern, then $\mathcal{T}_{n,k}(C)$ is a disjoint union of cycles. This is so because no d_i equals d_k and so all distinct permutations of the entries $d_1, \ldots, d_{k-1}$ can be partitioned into cyclic permutations of a fixed few. For example, with $C = [1, 1, 2, 2, 3; 12]$ we obtain the d-sets $(1, 1, 2, 2, 3; 12), (1, 1, 2, 3, 2; 12), (1, 1, 3, 2, 2; 12), (1, 2, 1, 2, 3; 12), (1, 2, 1, 3, 2; 12)$, and $(1, 3, 1, 2, 2; 12)$, along with all their five cyclic permutations each. Those of the last d-set induce the cycle of edges

$$\begin{aligned}((1,3,1,2,2)) &\to ((3,1,2,2,1)) \to ((1,2,2,1,3)) \\ &\to ((2,2,1,3,1)) \to ((2,1,3,1,2)) \to ((1,3,1,2,2))\end{aligned}$$

in $\mathcal{T}_{21,6}(C)$. An example of why this breaks down for bad d-patterns is the d-class $C = [1, 1, 4, 4]$ for 4-subsets of $[10]$, containing only the two d-sets $(1, 1, 4, 4)$ and $(1, 4, 1, 4)$. If we try to use the representation $[1, 1, 4; 4]$, then that induces the representations $(1, 1, 4; 4)$ and $(1, 4, 1; 4)$, which are the edges $((1, 1)) \to ((1, 4)) \to ((4, 1))$ in $\mathcal{T}_{10,4}(C)$. We can never form cycles this way.

We quickly see from this that if $\mathcal{T}_{n,k}$ is connected, then it is Eulerian, being a union of cycles. Two d-classes C_1 and C_2 being connected in $\mathcal{H}_{n,k}$ means that $\mathcal{T}_{n,k}(C_1)$ and $\mathcal{T}_{n,k}(C_2)$ share many of the same vertices. In particular, if C_1 and C_2 share $d_1, \ldots, d_{k-2}$, then their corresponding graphs share every permutation of them as vertices. For example, $((2, 2, 1, 3)), ((2, 2, 3, 1))$, and $((2, 1, 2, 3))$ are the shared vertices of $[2, 2, 1, 3, 6; 7]$ and $[1, 1, 2, 2, 3; 12]$ above. However, this does not mean that the union of the two graphs $\mathcal{T}_{n,k}(C_1)$ and $\mathcal{T}_{n,k}(C_2)$ is connected, although since $\mathcal{H}_{n,k}$ is connected, we do get paths to and from every d-set and $(1, 1, \ldots, 1)$, so that the union over all d-classes produces the connected graph $\mathcal{T}_{n,k}$. □

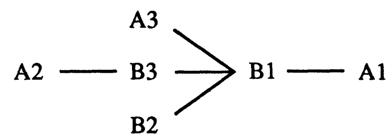

FIG. 2. *Connectedness of* $\mathcal{H}_{n,3}$.

LEMMA 8. *Let* $n_0(3) = 8, n_0(4) = 9,$ *and* $n_0(6) = 17$. *Then* $\mathcal{H}_{n,k}$ *is connected for* $k = 3, 4$ *and* 6 *with* $n \geq n_0(k)$ *and* $\gcd(n, k) = 1$.

Proof of Lemma 8.

Case k = 3. We assume $n \geq 8$ and break up the d-classes into the following types.

(A) d-pattern $\langle 2, 1 \rangle : [a, a; b]$
  (1) $a = 1$
  (2) $a > 1, b = 1$ (for n odd only)
  (3) $a > 1, b > 1$
(B) d-pattern $\langle 1, 1, 1 \rangle : [a, b; c]$
  (1) $1 = a < b < c$
  (2) $1 < a < b < c$
  (3) $2 = a < b - 1 = c$ (for n odd only)

We show that $\mathcal{H}_{n,3}$ is connected by displaying the path of edges from the d-classes of each type to the type A1 in Fig. 2. If n is odd, then A2 $\to$ B3 $\to$ B1 by $[a, a; 1] \to [2, a; a - 1] \to [1, 2; c]$. Then A3 $\to$ B1 by $[a, a; b] \to [1, a; c]$, B2 $\to$ B1 by $[a, b; c] \to [1, b; d]$, and B1 $\to$ A1 by $[1, b; c] \to [1, 1; d]$.

Case k = 4. We assume $n \geq 9$ and break up the d-classes into the following types.

(A) d-pattern $\langle 3, 1 \rangle : [a, a, a; b]$
  (1) $a = 1$
  (2) $a > 3, b = 1$
  (3) $a > 1, b > 1$
(B) d-pattern $\langle 2, 1, 1 \rangle : [a, a, b; c]$
  (1) $a = 1, 1 < b < c$
  (2) $a > 1, 1 = b < c$
  (3) $a > 1, 1 < b < c$
(C) d-pattern $\langle 1, 1, 1, 1 \rangle : [a, b, c; d]$
  (1) $1 = a < b < c < d$
  (2) $1 < a < b < c < d$

We show $\mathcal{H}_{n,4}$ is connected by displaying the path of edges from the d-classes of each type to the type A1 in Fig. 3. A2 $\to$ B3 $\to$ C1 by $[a, a, a; 1] \to [a, a, 2; a - 1] \to [1, 2, a; b]$, C2 $\to$ C1 $\to$ B1 by $[a, b, c; d] \to [1, b, c; e] \to [1, 1, c; f]$, A3 $\to$ B2 $\to$ B1 by $[a, a, a; b] \to [a, a, 1; c] \to [1, 1, a; d]$, and B1 $\to$ A1 by $[1, 1, a; b] \to [1, 1, 1; c]$.

Case $k = 6$.
We assume $n \geq 17$ and break up the d-classes into the following types.
(A) d-pattern $\langle 5, 1 \rangle : [a, a, a, a, a; b]$
  (1) $a = 1$
  (2) $a > 1, b > 1$
  (3) $a > 1, b = 1$
(B) d-pattern $\langle 4, 1, 1 \rangle : [a, a, a, a, b; c]$

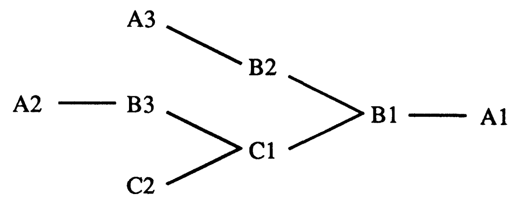

FIG. 3. *Connectedness of* $\mathcal{H}_{n,4}$.

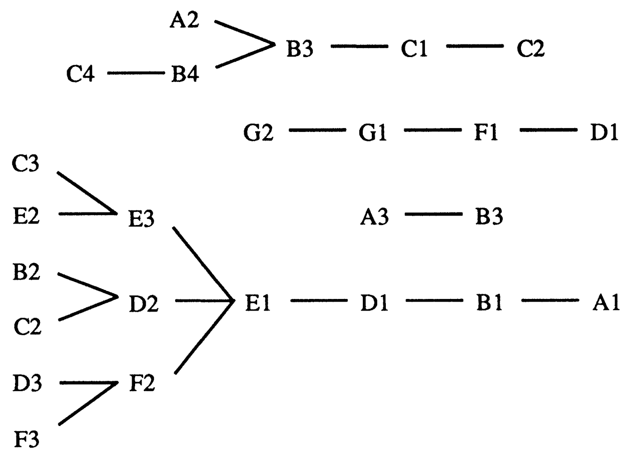

FIG. 4. *Connectedness of* $\mathcal{H}_{n,6}$.

(1) $a = 1, 1 < b < c$
(2) $a > 1, 1 < b < c - 2$
(3) $a > 1, 1 = b < c - 2$
(4) $a > 1, b = c + 1$

(C) d-pattern $\langle 3,2,1,1\rangle : [a,a,a,b,b;c]$
(1) $a > b$
(2) $a < b, c > 1$
(3) $2 < a < b, c = 1$
(4) $2 = a < b, c = 1$

(D) d-pattern $\langle 3,1,1,1\rangle : [a,a,a,b,c;d]$
(1) $a = 1, 1 < b < c < d$
(2) $a > 1, 1 = b < c < d$
(3) $a > 1, 1 < b < c < d$

(E) d-pattern $\langle 2,2,1,1\rangle : [a,a,b,b,c;d]$
(1) $1 = a < b, 1 < c < d$
(2) $1 < a < b, 1 < c < d$
(3) $1 < a < b, 1 = c < d$

(F) d-pattern $\langle 2,1,1,1,1\rangle : [a,a,b,c,d;e]$
(1) $a = 1, 1 < b < c < d < e$
(2) $a > 1, 1 = b < c < d < e$
(3) $a > 1, 1 < b < c < d < e$

(G) d-pattern $\langle 1,1,1,1,1,1\rangle : [a,b,c,d,e;f]$
(1) $1 = a < b < c < d < e < f$
(2) $1 < a < b < c < d < e < f$

We show $\mathcal{H}_{n,6}$ is connected by displaying the path of edges from the d-classes of each type to the type A1 in Fig. 4. C4 $\to$ B4 $\to$ B3 by $[2,2,2,b,b;1] \to [2,2,2,2,b;b-1] \to [2,2,2,2,1;c]$, A2 $\to$ B3 $\to$ C1 by $[a,a,a,a,a;b] \to [a,a,a,a,1;c] \to [a,a,a,1,$

$1; d]$, A3 $\to$ B2 by $[a, a, a, a, a; 1] \to [a, a, a, a, 2; a-1]$, and C1 $\to$ C2 $\to$ D2 by $[a, a, a, b, b; c] \to [b, b, b, a, a; d] \to [b, b, b, 1, a; e]$. B2 $\to$ D2 $\to$ E1 by $[a, a, a, a, b; c] \to [a, a, a, 1, b; d] \to [1, 1, a, a, b; e]$, C3 $\to$ E3 by $[a, a, a, b, b; 1] \to [a, a, b, b, 1; c]$, and E2 $\to$ E3 $\to$ E1 by $[a, a, b, b, c; d] \to [a, a, b, b, 1; e] \to [1, 1, a, a, b; f]$. D3 $\to$ F2 by $[a, a, a, b, c; d] \to [a, a, 1, b, d; e]$, F3 $\to$ F2 $\to$ E1 by $[a, a, b, c, d; e] \to [a, a, 1, c, d; f] \to [1, 1, a, a, d; g]$, and G2 $\to$ G1 $\to$ F1 by $[a, b, c, d, e; f] \to [1, b, c, d, e; g] \to [1, 1, c, d, e; h]$. And finally E1 $\to$ D1 by $[1, 1, b, b, c; d] \to [1, 1, 1, b, c; d]$, F1 $\to$ D1 by $[1, 1, b, c, d; e] \to [1, 1, 1, c, d; f]$, and D1 $\to$ B1 $\to$ A1 by $[1, 1, 1, b, c; d] \to [1, 1, 1, 1, b; e] \to [1, 1, 1, 1, 1; f]$. □

Proof of Theorem 4. The proof follows from Lemmas 5–8. □

We finally note that Jackson [J2] has recently found by computer Ucycles for the values of (n, k) equal to $(10, 4), (8, 5), (9, 5), (10, 6), (11, 6), (10, 7), (11, 7)$, and $(11, 8)$.

REFERENCES

[B] N. G. DE BRUIJN, *A combinatorial problem*, Nederl. Akad. Wetensch. Proc., 49 (1946), pp. 758–764.

[C] F. R. K. CHUNG, P. DIACONIS, AND R. I. GRAHAM, *Universal cycles for combinatorial structures*, Discrete Math., 110 (1992), pp. 43–59.

[F] H. FREDERICKSEN, *A survey of full length nonlinear shift register cycle algorithms*, SIAM Rev., 24 (1982), pp. 195–221.

[G] I. J. GOOD, *Normally recurring decimals*, J. London Math. Soc., 21 (1946), pp. 167–169.

[H] G. HURLBERT, *Universal Cycles: On beyond De Bruijn*, Ph.D. thesis, Rutgers University, New Brunswick, NJ, 1990.

[J1] B. JACKSON, *Universal cycles for k-subsets and k-permutations*, Discrete Math., 117 (1993), pp. 141–150.

[J2] B. JACKSON, personal communication May 1993.

[M] C. FLYE-SAINTE MARIE, *Solution to problem number* 58, l'Intermediaire des Mathematicieus, 1 (1894), pp. 107–110.