MULTICOVER UCYCLES

Glenn Hurlbert
Department of Mathematics
Arizona State University
Tempe, AZ 85287–1804

Abstract. A *Universal Cycle*, or *Ucycle*, for k-subsets of $[n] = \{1, 2, \ldots, n\}$ is a cyclic sequence of $\binom{n}{k}$ integers with the property that every k subset of $[n]$ appears exactly once consecutively in the sequence. A t*-cover Ucycle* for k-subsets of $[n]$ is a cyclic sequence of $t\binom{n}{k}$ integers with the property that every k-subset of $[n]$ appears exactly t times consecutively in the sequence. Here we investigate the minimal number $t = U(n, k)$ for which there is a t-cover Ucycle.

1. Introduction.

An example of a *de Bruijn Cycle* is the sequence 00011101. We say that it is binary of order 3 because, when thought of cyclically, each binary triple appears exactly once consecutively. Indeed, the sequence has 2^3 digits and the list of triples in the order they appear is 000, 001, 011, 111, 110, 101, 010, 100. It is well known that k-ary de Bruijn Cycles of order n exist for all k and n, and there are many interesting algorithms which are used to construct them. It is also known exactly how many there are (see [1], [3], [4], [5], [6]).

A generalization introduced by Chung, Diaconis, and Graham [2] is the *Universal Cycle*, or *Ucycle*, for k-subsets of $[n] = \{1, 2, \ldots, n\}$. It is a cyclic sequence of $\binom{n}{k}$ integers with the property that every k-subset of $[n]$ appears exactly once as a contiguous subsequence. The ambitious reader may verify that the following sequence is an example with $n = 8$ and $k = 3$ (for example, $\{2, 7, 8\}$ appears as the 21^{st} sequence 782 in Figure 1).

1356725 6823472 3578147 8245614 5712361 2467836 7134583 4681258

Figure 1.

A *multi-cover Ucycle* for k-subsets of $[n]$, introduced in [6], is a cyclic sequence of integers with the property that each k-subset of $[n]$ appears consecutively equally often. If t is the number of occurrences of each k-subset, then we say it is a *t-cover Ucycle*. Thus, a Ucycle is a 1-cover Ucycle. Here we investigate the function $U(n, k)$ which is the minimal t for which a t-cover Ucycle exists.

For example with $k = 2$ an Eulerian circuit in the complete graph K_n (n odd) produces a Ucycle ($n = 5$: 1234513524) so that $U(2m + 1, 2) = 1$. However, K_n, n even, has no Eulerian circuit while any Eulerian circuit in the graph obtained from K_n by duplicating its edges forms a double-cover Ucycle ($n = 6$: 123456123456135141363524624625) so that $U(2m, 2) = 2$.

We remark that the sequence $12 \cdots n$ is a Ucycle for $k = 1$ and for $n = k+1$. Such trivial examples will not be considered here so we will always assume that $k > 1$ and $n > k + 1$.

2. Theorems.

We begin with a simple necessary condition.

Fact 1. *If there is a t-cover Ucycle for k-subsets of $[n]$, then n divides $t\binom{n}{k}$.*

Proof of Fact 1. A t-cover Ucycle has length $t\binom{n}{k}$ and each integer occurs equally often.$\square$

For $t = 1$ this necessary condition is conjectured to be sufficient for n large enough in [2]. We believe the same should be true for general t.

Theorem 2. *If $k = 2, 3, 4,$ or 6, $gcd(n, k) = 1$, and $n \geq n_0(k)$ then $U(n, k) = 1$.*

The proof is to be found in [7] (see also [8]) with values of $n_0(2) = 5$, $n_0(3) = 8$, $n_0(4) = 9$, and $n_0(6) = 17$.

In this paper we show that $U(n, k)$ is well-defined (that is, $\forall n \geq k \exists t \ldots$) and in addition we present some specific values.

Theorem 3. *$U(n, k) \leq k$ for all $n \geq k$.*

Corollary 4. *If k divides n, then $U(n, k) = k$.*

We will postpone the proof of Theorem 3 until section 5 and proceed with the proof of Corollary 4. We will use a very convenient fact (see [10], [11]) which gives the exact power of primes dividing binomial coefficients. Given a prime p and integer x we let $\nu_p(x)$ be the maximum exponent r such that p^r divides x.

Fact 5 (Kummer's Theorem). $\nu_p(\binom{n}{k})$ equals the number of borrows needed when subtracting k from n in base p.

For example, $\binom{22}{3} = 2^2 \cdot 5 \cdot 7 \cdot 11$. 22 base 2 is 10110, 3 base 2 is 11, and 2 borrows are needed. But no borrows are needed in base 3 $(211 - 10)$.

Proof of Corollary 4. Suppose $n = mk$ and let $t \leq k$. If we assume that n divides $t\binom{n}{k}$ then we must show that $t = k$. Equivalently, we will assume that k divides $t\binom{n-1}{k-1}$ and show that k divides t. Let $k = \Pi p_i^{q_i}$. Then $n - 1 = mk - 1 = m\Pi p_i^{q_i} - 1$ which, in base p_i, ends in q_i $(p_i - 1)$'s. Likewise, $k - 1$ ends in q_i $(p_i - 1)$'s and so subtracting $(k - 1)$ from $(n - 1)$ in base p_i requires no borrowing. Hence, by Fact 5, p_i does not divide $\binom{n-1}{k-1}$, implying that $p_i^{q_i}$ divides t for all i. Thus $t = k$. $\square$

Theorem 2 and Corollary 4 combine to yield all values of $U(n, 2)$ for $n \geq 4$, all values of $U(n, 3)$ for $n \geq 6$, and all values of $U(n, 4)$ for all $n \geq 8$ except $n \equiv 2 \bmod 4$. In this case we prove

Theorem 6. $U(n, 4) \leq 2$ *if* $n \equiv 2 \bmod 4$.

In light of Fact 1 we see that $U(n, 4) = 2$ for $n \equiv 6 \bmod 8$. The remaining mystery is whether $U(n, 4) = 1$ or 2 for $n \equiv 2 \bmod 8$, although Brad Jackson [9] has recently discovered that $U(10, 4) = 1$.

Theorem 7. *For n relatively prime to 5 we have $U(n, 5) \leq 2$, provided that $n = 12$ or $n \geq 16$ and $n \not\equiv 2 \bmod 3$.*

The proofs of Theorems 6 and 7 will also be postponed until Section 5. Let us first remark that our necessary condition is satisfied in these cases. Indeed, one can see that if $\gcd(n, k) = t$, then n divides $t\binom{n}{k}$ (so that n divides $k\binom{n}{k}$ always holds—we need this for

Theorem 3). In this case $n = mt$ and $k = ht$ with gcd $(m, h) = 1$, and $\binom{n-1}{k-1} = \frac{k}{n}\binom{n}{k} = \frac{h}{m}\binom{n}{k}$ is an integer, so that m must divide $\binom{n}{k}$. Thus, n divides $t\binom{n}{k}$. We also note that $U(9,5) = 1$ [9].

3. Notation and Terminology.

Let $S = \{s_1, \dots, s_k\}$, $s_i < s_{i+1}$, be a k-subset of $[n]$. Define its *difference set*, or *d-set*, $D(s) = (d_1, \dots, d_k)$ by $d_i = s_{i+1} - s_i$, where indices are modulo k and arithmetic in modulo n. (Keep in mind that D is an ordered set.) Two d-sets are *equivalent* if one is a cyclic permutation of the other. Thus, we say that any rotation of S belongs to the same d-set as S, where a *rotation* of S is any set $S + r = \{s_1 + r, \dots, s_k + r\}$ with r an integer. Two d-sets will belong to the same *d-class* whenever one is any permutation of the other. The reader should notice that the family of all d-classes is simply the family of all unordered partitions of n (not $[n]$) 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], $\dots$, 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 $\{1,3,5,22,24\}$, $\{2,4,6,23,25\}, \dots$, 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.

Now given the d-set $(d_1, \dots, d_{k-1}, d_k)$ we distinguish one of its coordinates (which we may assume to be d_k because of cyclic permutations) in the representation $(d_1, \dots, d_{k-1}; d_k)$

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 will be distinguished (unused) in each of its d-sets. Thus, to avoid ambiguity, it is important that d_k be unique, or in other words, that the corresponding d-pattern is 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 $\{11,2,3,5\}$.

We will need to define two graphs, the *transition (di)graph* $\mathcal{T}_{n,k}$ and the *class graph* $\mathcal{H}_{n,k}$, each depending on the chosen representations of the d-classes. 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 the vertices of $\mathcal{T}_{n,k}$, whose directed edges are precisely the representations involved.

Before introducing $\mathcal{H}_{n,k}$ let's get 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 which are representations of d-sets belonging to C. We now let the vertices of $\mathcal{H}_{n,k}$ be all possible d-classes and let the undirected edges of $\mathcal{H}_{n,k}$ 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 $\mathcal{H}_{21,6}$.

4. An Example.

The proof of Theorem 2 essentially rests on two lemmas (see [7]).

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

Lemma B. *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.*

To give an example of the role of these two lemmas in Ucycle construction, let $n = 8$ and $k = 3$. The only d-patterns are $\langle 1,1,1 \rangle$ and $\langle 2,1 \rangle$, and both are good. The reader might enjoy listing all d-classes and d-sets, finding "nice" representations for them, and comparing their results with the graphs $\mathcal{T}_{8,3}$ and $\mathcal{H}_{8,3}$ in figures 2 and 3, respectively.

Figure 2. $\mathcal{T}_{8,3}$

Figure 3. $\mathcal{H}_{8,3}$

The Eulerian circuit 2211331 in figure 2 corresponds to a listing of all d-sets and produces the differences in the first block, 1356725, along with the first digit of the next block from figure 1. Since the *circuit 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 [9] occurs exactly once! That 5 is relatively prime to 8 turns out to be an unnecessary luxury (see [7]). For example, if the sum turned out to be $2 \bmod 8$ then we would produce two disjoint cycles, one beginning (without loss of generality) with $123\cdots x$ and the other with $234\cdots y$. We can then join these two cycles as $1234\cdots y23\cdots x$. This is part of a general method we call *insertion*.

5. Proofs.

The proof of Theorem 3 is simple.

Proof of Theorem 3. We show the existence of a k-cover Ucycle for k-subsets of $[n]$ for any $n \geq k$. For each d-set $(d_1, \ldots, d_k)$ we use every possible distinct representation $(d_1, \ldots, d_{k-1}; d_k), (d_2, \ldots, d_k; d_1), \ldots$ and $(d_k, \ldots, d_{k-2}; d_{k-1})$. Now build the transition graph $\mathcal{T}'_{n,k}$ on all prefixes and suffixes as vertices with the obvious directed edges. This graph is the union of directed cycles (one per d-set) and so is Eulerian if connected. But this is easy to show since every representation exists. We simply find a path from vertex $((d_1, \ldots, d_{k-2}))$ to vertex $((1, \ldots, 1))$ by sequentially changing each d_i to a 1. Every intermediate vertex $((d_{j+1}, \ldots, d_{k-2}, 1, \ldots, 1))$ exists since, if $s_j = n - [d_j + d_{j+1} + \cdots + d_{k-2} + 1 + \cdots + 1]$ then the d-set $(d_j, d_{j+1}, \ldots, d_{k-2}, 1, \ldots, 1, s_j)$ has $(d_j, d_{j+1}, \ldots, d_{k-2}, 1, \ldots, 1; s_j)$ as one of its representations. This representation corresponds to the edge from $((d_j, d_{j+1}, \ldots, d_{k-2}, 1, \ldots, 1))$ to $((d_{j+1}, \ldots, d_{k-2}, 1, \ldots, 1))$. Any Eulerian circuit gives rise to n disjoint cycles (since the circuit sum is $0 \bmod n$) which n insertions join together. Since for every set each of its k

representations is used, each set appears k times. $\square$

Proof of Theorem 6. We prove this by showing the existence of a double cover. Let $\mathcal{T}'_n$ be the usual transition graph, except in this case we only allow the vertices to be prefixes and suffixes of d-sets belonging to good d-patterns. As in the other cases with $k = 4$ (see [8]), the graph is connected and Eulerian, so we can construct (using insertions if necessary) a semi-UCycle, one which covers once all sets but those from bad d-patterns. In fact, we attach a duplicate copy so as to double-cover these good sets. Call this sequence C_0.

What are the bad sets we are missing? The only bad d-pattern is $\langle 2, 2\rangle$ so the only bad d-classes are of the form $[x, x, y, y]$. These contain the d-sets (x, x, y, y) (type A) and (x, y, x, y) (type B). Let $x < y$ and define the graph $\mathcal{T}^*_n(x)$ to be the cycle $((x, x)) \to ((x, y)) \to ((y, x)) \to ((x, x))$. This yields representations $[x, x, y; y]$ of A, $[x, y, x; y]$ of B, and $[y, x, x; y]$ of A again. Of course, this is exactly what we want to produce a double-cover, the two representations of A being obvious, the one of B more subtle. We notice the symmetry of type B sets, those starting at i being identical to those starting at $i + n/2$, and then realize that this is what produces the double-cover. As for the technical details, let $s = 2x + y$ and $r = gcd(s, n)$. Then by tracing $\mathcal{T}^*_n(x)$ repeatedly we produce sets $R, S, T, R{+}s, S{+}s, T{+}s, R{+}2s, S{+}2s, T{+}2s, \ldots,$ and $R{-}s, S{-}s, T{-}s$, and call this sequence $C_1(x)$. Of course, we really produce the disjoint sequences $C_1(x), \ldots, C_r(x)$. Because $((x, x))$ is a vertex in $\mathcal{T}'_n$, we can insert each of the cycles $C_i(x)$ into C_0, thus producing our desired double-cover. That $((x, x))$ is a vertex is witnessed by the representation $[x, x, y - 1; y + 1]$ which always exists since $x < y$ so that $y > 1$. $\square$

In fact, one can realize how close we actually come to producing a UCycle for 4-subsets of $[n]$ with $n \equiv 2 \bmod 8$. Let C_0 be a single cover of the good sets and consider the graphs

$\mathcal{T}_n^*(x)$ with x even rather than $x < y$. Thus, y is odd. From the cycles in $\mathcal{T}_n^*(x)$ we can produce a single cover of all the bad sets following the example below with $n = 10$ and $x = 2$. We cannot repeat the cycle twice for fear of double-covering, and instead form $5 = n/2$ disjoint cycles.

$\underline{[2,2,3;3]}$	$\underline{[2,3,2;3]}$	$\underline{[3,2,2;3]}$
$\{1,3,5,8\}$	$\{3,5,8,0\}$	$\{5,8,0,2\}$
$\{3,5,7,0\}$	$\{5,7,0,2\}$	$\{7,0,2,4\}$
$\{5,7,9,2\}$	$\{7,9,2,4\}$	$\{9,2,4,6\}$
$\{7,9,1,4\}$	$\{9,1,4,6\}$	$\{1,4,6,8\}$
$\{9,1,3,6\}$	$\{1,3,6,8\}$	$\{3,6,8,0\}$

The first representations of type A sets contains those which begin with an odd number and the second, after we rotate the elements of each set so as to look like the first (for example $\{5, 8, 0, 2\}$ becomes $\{8, 0, 2, 5\}$), contains those which begin with an even number. This shows the importance of making y odd. So certainly we have single-covered all the bad sets and this is always possible for such n. Our new problem is that all but one of the insertions work! That is, we can always insert at vertex $((x, x))$ as before except in the case that $y = 1$. Here, $[x, x, y - 1; y + 1]$ does not exist so we fail to be able to insert. But oh, so close!

Finally, on to our last result.

Proof of Theorem 7. We employ the same methods as in the previous theorem. Define $\mathcal{T}_n'$ to have vertices all prefixes and suffixes of good d-sets; those belonging to one of the good d-patterns $\langle 4, 1\rangle$, $\langle 3, 1, 1\rangle$, $\langle 2, 2, 1\rangle$, $\langle 2, 1, 1, 1\rangle$, or $\langle 1, 1, 1, 1, 1\rangle$. Again, $\mathcal{T}_n'$ is clearly Eulerian provided it is connected, which we show by proving the connectedness of its corresponding class graph $\mathcal{H}_n'$. We assume $n = 12$ or $n \geq 16$ and $n \not\equiv 2 \bmod 3$ (and not divisible by 5) and break up the d-classes into the following types.

$A)$ d-pattern $\langle 4,1\rangle : [a,a,a,a;b]$

1) $a = 1$

2) $a > 1,\ \ b = 1$

3) $a > 1,\ \ b > 1$

$B)$ d-pattern $\langle 3,1,1\rangle : [a,a,a,b;c]$

1) $a = 1$

2) $a < 1, 1 = b < c$

3) $a = 2 < b < c - 1$

4) $a = 2 < b = c + 1$

5) $a > 2, b = a + 1, c = 2$

6) $a > 2,$ other $b < c$

$C)$ d-pattern $\langle 2,2,1\rangle : [a,a,b,b;c]$

1) $1 = a < b$

2) $1 < a < b, c > 1$

3) $2 = a < b, c = 1$

4) $2 < a < b - 1, c = 1$

5) $2 < a = b - 1, c = 1$

$D)$ d-pattern $\langle 4,1\rangle : [a,a,b,c;d]$

1) $a=1, 1<b<c<d$

2) $a>1, 1=b<c<d$

3) $a>1, 2=b<c=d+1$

4) $a>1, 2=b<c<d-1$

5) $a>1, 2<b<c<d$

$E)$ d-pattern $\langle 4,1\rangle : [a,b,c,d;e]$

1) $a=1$

2) $a>1$

Figure 4. Connectedness of $\mathcal{H}'_n$.

We show $\mathcal{H}'_n$ is connected by displaying the path of edges from the d-classes of each type to the type A1 in figure 4. C3 $\rightarrow$ B4 $\rightarrow$ B2 by $[2,2,b,b;1] \rightarrow [2,2,2,b;b-1] \rightarrow [2,2,2,1;c]$, C5 $\rightarrow$ B5 $\rightarrow$ B2 by $[a,a,a+1,a+1;1] \rightarrow [a,a,a,a+1;2] \rightarrow [a,a,a,1;a+2]$, B3 $\rightarrow$ B2 $\rightarrow$ C1 by $[2,2,2,b;c] \rightarrow [2,2,2,1;d] \rightarrow [1,1,2,2;e]$, and A3 $\rightarrow$ B2 by $[a,a,a,a;b] \rightarrow [a,a,a,1;c]$. C4 $\rightarrow$ D3 $\rightarrow$ E1 by $[a,a,b,b;1] \rightarrow [a,a,2,b;b-1] \rightarrow [1,2,a,b;c]$, D4 $\rightarrow$

E1 by $[a, a, 2, c; d] \to [1, 2, a, c; e]$, D5 $\to$ E1 by $[a, a, b, c; d] \to [1, a, b, c; e]$, and E2 $\to$ E1 by $[a, b, c, d; e] \to [1, b, c, d; f]$. A2 $\to$ B6 $\to$ D2 by $[a, a, a, a; 1] \to [a, a, a, 2; a-1] \to [a, a, 1, 2; b]$, C2 $\to$ D2 $\to$ D1 by $[a, a, b, b; c] \to [a, a, 1, b; d] \to [1, 1, a, b; e]$, E1 $\to$ D1 $\to$ B1 by $[1, b, c, d, e] \to [1, 1, c, d; f] \to [1, 1, 1, d; g]$, and C1 $\to$ B1 $\to$ A1 by $[1, 1, b, b; c] \to [1, 1, 1, b; d] \to [1, 1, 1, 1; e]$.

Now that $\mathcal{T}_n'$ is Eulerian, we produce our semi-UCycle as usual, using insertions as necessary. In fact, we again wish to double-cover our good subsets and call the resulting sequence C_0. Next we must add the bad subsets.

Since our only bad d-pattern is $\langle 3, 2 \rangle$, we are missing those sets whose d-class is $[x, x, x, y, y]$. These include the two d-sets (x, x, x, y, y) (type A) and (x, x, y, x, y) (type B). We now form the graph $\mathcal{T}_n^*(x)$ as the cycle $((x, x, x)) \to ((x, x, y)) \to ((x, y, x)) \to ((y, x, x)) \to ((x, x, x))$ so as to produce the representations $(x, x, x, y; y)$, $(x, x, y, x; y)$, $(x, y, x, x; y)$, and $(y, x, x, x; y)$. Each type is now double-covered and we produce our disjoint cycles as in the proof of the previous theorem. Insertions are always possible, as witnessed by the existence of the d-set representation $(x, x, x, y-1; y+1)$, which exists because $y \neq 1$ (else $n \equiv 2 \bmod 3$). $\square$

6. References

[1] T. van Ardenne-Ehrenfert, N.G. de Bruijn, Circuits and trees in ordered linear graphs, Simon Steven **28** (1951), 203–217.

[2] F.R.K. Chung, P. Diaconis, R.I. Graham, Universal Cycles for Combinatorial Structures, Discrete Math., to appear.

[3] N.G. de Bruijn, A combinatorial problem, Nederl. Akad. Wetensch, Proc. 49 (1946), 758–764.

[4] C. Flye-Sainte Marie, Solution to problem number 58, l'Intermediaire des Mathematiciens **1** (1894), 107–110.

[5] H. Fredericksen, A survey of full length nonlinear shift register cycle algorithms, SIAM Review **24**, 2 (1982), 195–221.

[6] G. Hurlbert, Ph.D. thesis, 1990.

[7] G. Hurlbert, On Universal Cycles, submitted to SIAM J. Disc. Math.

[8] B. Jackson, Universal Cycles for k-subsets and k-permutations, Discrete Math., to appear.

[9] B. Jackson, personal communication.

[10] E.E. Kummer, Über de Ergänz ungssätze zu den allgemeinen Reziprozitätsgesetzen, J. Reine u. Angew. Math. **44** (1852), 93–146.

[11] D. Singmaster, Notes on binomial coefficients I—A generalization of Lucas' Congruence, J. London Math. Soc. (2), **8** (1974), 545–548.