SIAM J. DISCRETE MATH.
Vol. 23, No. 3, pp. 1441–1449
© 2009 Society for Industrial and Applied Mathematics

Downloaded 03/14/23 to 129.219.247.33 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

NEAR-UNIVERSAL CYCLES FOR SUBSETS EXIST*

DAWN CURTIS†, TAYLOR HINES†, GLENN HURLBERT†, AND TATIANA MOYER†

Abstract. Let S be a cyclic n-ary sequence. We say that S is a *universal cycle* ((n,k)*-Ucycle*) for k-subsets of $[n]$ if every such subset appears exactly once contiguously in S, and is a *Ucycle packing* if every such subset appears at most once. Few examples of Ucycles are known to exist, so the relaxation to packings merits investigation. A family $\{S_n\}$ of (n,k)-Ucycle packings for fixed k is a *near-Ucycle* if the length of S_n is $(1-o(1))\binom{n}{k}$. In this paper we prove that near-(n,k)-Ucycles exist for all k.

Key words. universal cycle, Ucycle, packing, covering

AMS subject classifications. Primary, 05B30; Secondary, 05C35, 05C45, 68R15

DOI. 10.1137/080735795

1. Introduction. A *universal cycle (Ucycle) for k-subsets of* $[n]$, denoted (n,k)-UCS, is a cyclic sequence of integers from $[n]=\{1,2,\ldots,n\}$ with the property that each k-subset of $[n]$ appears exactly once consecutively in the sequence. For example, 1234524135 is a universal cycle for pairs of [5]. Chung, Diaconis, and Graham [1] defined Ucycles for a general class of combinatorial structures, generalizing both de Bruijn sequences and Gray codes (see [7]). The books [5, 6] contain a wealth of information about generating such things efficiently.

A necessary condition for the existence of an (n,k)-UCS is that $n \mid \binom{n}{k}$. This is because symmetry demands that each symbol appear equally often. The authors of [1] conjectured that the necessary condition is sufficient for large enough n in terms of k (some evidence [4] suggests $n \geq k+3$) and offered \$100 for its proof.[1]

CONJECTURE 1 (see [1]). *For all $k \geq 2$ there exists $n_0(k)$ such that for $n \geq n_0(k)$ Ucycles for k-subsets of $[n]$ exist if and only if n divides $\binom{n}{k}$.*

Progress on the conjecture has been slow. The $k=2$ case is trivial, corresponding to the existence of Eulerian circuits in K_n if and only if n is odd. Jackson [3] proved the conjecture for $k=3$ and constructed ucyles for $k=4$ and odd n, leaving the case $n \equiv 2 \bmod 8$ unresolved. In [2] we find the following result.

RESULT 2 (see [2]). *Let $n_0(3)=8$, $n_0(4)=9$, and $n_0(6)=17$. Then (n,k)-UCs exist for $k=3$, 4, and 6 with $n \geq n_0(k)$ and $\gcd(n,k)=1$.*

Note that $n_0(k)=3k$ suffices for $k \in \{3,4,6\}$. It would be nice to lower $3k$ as much as possible. To this end, Stevens et al. [8] proved the following.

RESULT 3 (see [8]). *No $(k+2,k)$-UC exists for $k \geq 2$.*

Combined with particular computer examples found by Jackson [4] (e.g., $(n,k)=(10,4)$), this suggests that $n_0(k)=k+3$ may suffice.

In cases where no Ucycles have been found, or none exist, it is natural to look for cycles with as many distinct subsets as possible, that is, a *Ucycle packing*, such as

*Received by the editors September 19, 2008; accepted for publication (in revised form) July 14, 2009; published electronically September 23, 2009.
http://www.siam.org/journals/sidma/23-3/73579.html
†Department of Mathematics and Statistics, Arizona State University, Tempe, AZ 85287-1804 (dawn.curtis@asu.edu, taylor.hines@asu.edu, hurlbert@asu.edu, tatiana.moyer@asu.edu).
[1]At the 2004 Banff Workshop on Generalizations of de Bruijn Cycles and Gray Codes it was suggested by the second author that a modest inflationary rate should revalue the prize near \$250.04.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/14/23 to 129.219.247.33 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

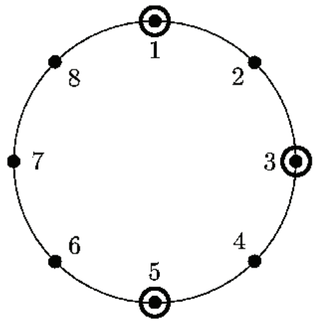


FIG. 1. *Form visualization of* $\{1, 3, 5\}$.

the sequence

$$S = 1345682\ 4678135\ 7812468\ 2345713\ 5678246\ 8123571\ 3456824\ 6781357$$

for $(n, k) = (8, 4)$. Note that no $(8, 4)$-UCS exists, and that S accounts for 56 of the possible 70 subsets. One might notice that these blocks shift upward by 3 mod 8 from one to the next. This is important in the techniques that follow.

In their paper, Stevens et al. show that the longest possible packing of a $(k+2, k)$-UCS has length $k+2$ and that a packing achieving this bound always exists. Compared to the potential $\binom{k+2}{k}$ length, this shows that for $n = k + 2$, we cannot get close to a full universal cycle. To establish this notion formally, we define a *near-Ucycle* packing as a sequence of Ucycle packings, one for each n, such that as $n \to \infty$, asymptotically few k-subsets are omitted from the (n, k)-UCS packing. That is, the length of cycle S_n is $(1 - o(1))\binom{n}{k}$. For example, if n is even and M is any perfect matching in K_n, then $K_n - M$ is Eulerian. In particular, any Eulerian circuit is a near-(n,2)-UCS of length $(1 - \frac{1}{n-1})\binom{n}{2}$. The purpose of this paper is to prove that near-Ucycles exist for all k.

THEOREM 4. *For all k, near-(n, k)-Ucycles exist.*

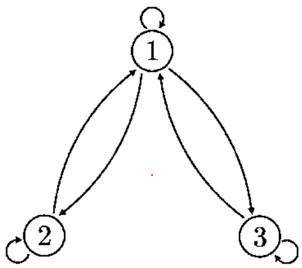
We proceed in the proof of this theorem by analyzing the construction of Ucycles. As we will show, we can create a Ucycle packing by selecting only those subsets that avoid certain structure. We prove that there are asymptotically few such structured subsets.

2. General technique. The general technique used to construct Ucycles originated from [3]. It consists of classifying the component subsets by their structure and ordering them accordingly. We write the k-subset S of $[n]$ as $S = \{s_1, \ldots, s_k\}$, with $s_i < s_{i+1}$, and define the *form* of S as $F = (f_1, \ldots, f_k)$ by $f_i = s_{i+1} - s_i$, where indices are modulo k and arithmetic is modulo n. That is, the form of a set is the ordered collection of distances between set elements (see Figure 1).

Consider the following example. For $(n, k) = (8, 3)$, the set $\{1, 3, 5\}$ has form $(2, 2, 4)$. We consider the cyclic permutations of a form to be equivalent, so $(2, 4, 2)$ and $(4, 2, 2)$ can both serve as the form of $\{1, 3, 5\}$. The choice of form has to do with how the form appears in the cycle. For example, the form $F = (4, 2, 2)$ makes the sets $\{1, 3, 5\}, \{2, 4, 6\}, \ldots,$ and $\{8, 4, 2\}$ appear as $513, 624, \ldots,$ and 482, respectively. Note that the last 2 in F is not used for these sets, and so we may represent F as $(4, 2; 2)$ or, more simply, $(4, 2)$. In our techniques below, we will restrict our attention to forms $(f_1, \ldots, f_{k-1}; f_k)$ having unique f_k. In fact, we choose f_k to be the largest unique entry.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/14/23 to 129.219.247.33 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy



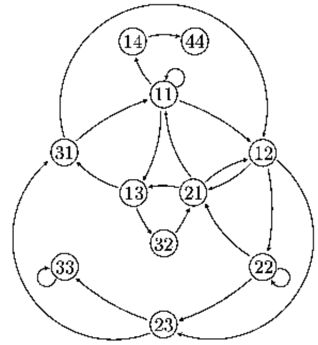


FIG. 2. *Transition graphs.*

It is important to note that the method used here to construct the forms is somewhat limited. As part of our definition, every form has entries whose sum is n. We call such forms *simple*. However, it is possible to relax this condition. For example, the set $\{5, 3, 1\}$ would have the form $(6, 6, 4)$. If we allow the sum of form entries to be a multiple of n, we would have more freedom in representing the subsets, and hence a better method of constructing Ucycles could be developed. These forms we define as *crossing*. In this paper, we will use only simple forms.

The purpose of the forms is to model what occurs in a Ucycle, namely, that if $s_0s_1 \ldots s_k$ appears in a Ucycle, then the forms $(f_1, \ldots, f_{k-1})$ and $(f_2, \ldots, f_k)$ of the sets $\{s_0, \ldots, s_{k-1}\}$ and $\{s_1, \ldots, s_k\}$ overlap on $(f_2, \ldots, f_{k-1})$. This motivates the following definitions.

For a given form representation $(f_1, \ldots, f_{k-1})$, we define $(f_1, \ldots, f_{k-2})$ to be its *prefix*, and $(f_2, \ldots, f_{k-1})$ to be its *suffix*.

The *transition graph*, denoted $\mathcal{T}_{n,k}$, is a graph whose vertices are the prefixes (and suffixes) of the form representations of (n, k)-UCS. The directed edges are the form representations, drawn from prefix to suffix. In order for this construction to produce a Ucycle, it is necessary that this transition graph be Eulerian.

Consider the cases of $(n, k) = (8, 3)$ and $(n, k) = (10, 4)$ and the forms of each. We use these forms to construct $\mathcal{T}_{8,3}$ and $\mathcal{T}_{10,4}$, as shown in Figure 2.

The condition of evenness is highlighted by these two examples. We know that a Ucycle is possible for $(n, k) = (8, 3)$ but not for $(n, k) = (10, 4)$. As shown in the following result, this is directly connected to the fact that $\mathcal{T}_{8,3}$ is Eulerian and $\mathcal{T}_{10,4}$ is not.

RESULT 5 (see [2]). *If $\mathcal{T}_{n,k}$ is Eulerian for some choice of form representations, then an (n, k)-UCS exists.*

As we can see in $\mathcal{T}_{10,4}$, the vertex 44 has no out degree. This is not necessarily the case. If we write the form $(1, 1, 4; 4)$ as $(4, 4, 1; 1)$, then we would instead connect $44 \to 41$. Since this form has no unique entry, this is left ambiguous. We call such forms *bad*. We define all forms with at least one unique element, a clear representative, as *good*.

Consider the previous example. We can ignore all bad forms of $(n, k) = (10, 4)$,

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

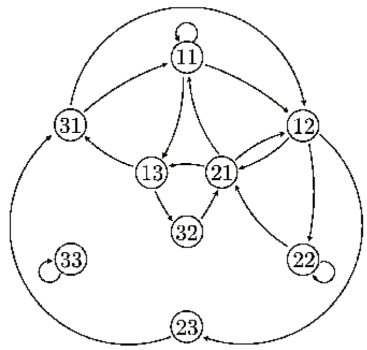


FIG. 3. *Good* $\mathcal{T}_{10,4}$.

and construct the transition graph, as shown in Figure 3.

In this case, ignoring the bad forms yields an even transition graph. As stated in the following result, this is true for any connected transition graph. Note, however, that it is not Eulerian since vertex ((33)) is isolated. This problem disappears if we also ignore all such isolated cycles, as stated in the following result. As we will show, the number of such isolated cycles is negligible as $n \to \infty$.

RESULT 6 (see [2]). *If* $\mathcal{H}_{n,k}$ *is connected and there are no bad forms for* k*-subsets of* $[n]$*, then* $\mathcal{T}_{n,k}$ *is Eulerian.*

We can construct a Ucycle packing by ignoring such bad forms for any (n, k) pair and restricting our attention to the largest component of $\mathcal{T}_{n,k}$, the component containing vertex $((1\ldots1))$.

3. Main result. In order to prove that a near-Ucycle packing is possible for any k, we must show that an asymptotically large proportion of sets can be included in a Ucycle. As we will show, the *good sets*, those belonging to good forms, can easily be included in a Ucycle packing, while the remaining *bad sets* cannot. As we have indicated, we will include only good sets. Of course, not all good sets can be included. It therefore remains to show that the good sets which can be included asymptotically outnumber all other sets, and that a Ucycle packing that includes each of these sets can be created.

3.1. Counting good sets. Consider the example of the $(8, 4)$-UCS. We know that each form must have four entries, and, since each form is simple, the sum of these entries must be eight. The number of times each entry appears is also important, since the bad forms will have no entry appearing only once. Therefore, we do not need all of its entries in order to determine whether or not a form is bad. We need only know how many times each entry appears. For example, the form $(1, 2, 2; 3)$ has the unordered *pattern* $\langle 1, 2, 1\rangle = \langle 2, 1, 1\rangle$, essentially the list of multiplicities.[2] Since this pattern contains a 1 as an entry, it has a unique element, and therefore all forms with this *good pattern* are good. In this case, the pattern entries define a partition of 4, since each form has 4 entries. In general, every pattern of an (n, k)-UCS is a partition of k.

Consider a *bad pattern* $P = \langle p_1, \ldots, p_t\rangle$. Clearly, $p_i \geq 2$ for all $1 \leq i \leq t$, so there exists a corresponding good pattern $P' = \langle p_1, \ldots, p_t - 1, 1\rangle$. We define the function

[2] We can also denote the form $(1, 2, 2, 3)$ as $(3^1 2^2 1^1)$, a notation we will find useful later.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/14/23 to 129.219.247.33 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Downloaded 03/14/23 to 129.219.247.33 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

$\phi : \Gamma \to \beta$ from the set of of good patterns to the set of bad patterns as $\pi(P) = P'$. Since this map applies to any bad pattern, we can see that $|\beta| \le |\Gamma|$.[3]

As we can see, many forms may belong to the same pattern. In fact, the forms belonging to a particular pattern satisfy the equation $\sum_{i=1}^{t} p_i x_i = n$ for some positive integers $x_1, \dots, x_t$, where $\langle p_1, p_2, \dots, p_t\rangle$ are the pattern entries. However, this equation imposes no order on the solution and thus does not distinguish between two different forms that share the same entries. We say that two such forms belong to the same *class*.

For example, the forms $(1, 2, 2; 3)$, $(2, 1, 2; 3)$, and $(2, 2, 1; 3)$ both share the class $[1, 2, 2; 3]$. By convention, we order the entries of a class from smallest to largest entry.

As we know, the classes belonging to a pattern P are all representations of n as a positive distinct integer linear combination of $p_1, \dots, p_t$. We count the classes with the aid of Schur's theorem. For $P = \langle p_1, \dots, p_t\rangle$, define the function

$$\psi(P) = \left|\left\{(x_1, \dots, x_t) \mid \sum_{j=1}^{t} p_j x_j = n, x_j \ge 0\right\}\right|.$$

THEOREM 7 (see [9]). *Suppose that* $\gcd(p_1, \dots, p_t) = 1$, *and define* $k = \sum_{j=1}^{t} p_j$ *and* $Q = \prod_{j=1}^{t} p_j$. *Then*

$$\psi(P) \sim \frac{n^{t-1}}{(t-1)!Q}.$$

Counting the classes requires a slightly more general result. Namely, we want to extend Schur's theorem as follows. We define

$$\psi'(P) = \left|\left\{(x_1, \dots, x_t) \mid \sum_{j=1}^{t} p_j x_j = n,\ x_j \ge 1,\ x_j \ne x_i\ \forall i \ne j\right\}\right|.$$

LEMMA 8. *Define* $g = \gcd(p_1, \dots, p_t)$, $k = \sum_{j=1}^{t} p_j$, *and* $Q = \prod_{j=1}^{t} p_j$. *Then*

$$\psi'(P) \sim \frac{(n-k)^{t-1}}{(t-1)!Qg^{t-1}} - \frac{(n-k)^{t-2}}{(t-2)!Qg^{t-2}} \sum_{i<j} \frac{p_i p_j}{p_i + p_j}.$$

Proof. Since all form entries are positive integers, we count only integer solutions to $\sum_{j=1}^{t} p_j x_j = n$ such that $x_j \ge 1$. This is equivalent to finding all integer solutions to the equation $\sum_{j=1}^{t} p_j(x_j + 1) = \sum_{j=1}^{t} p_j x_j + k = n$, without the positivity constraint.

We further modify the system to act for the condition $\gcd(p_1, \dots, p_t) = g$. The equation $\sum_{j=1}^{t} \frac{p_j}{g} x_j = \sum_{j=1}^{t} q_j x_j = \frac{n-k}{g}$ is the same as our original equation, and clearly $\gcd(q_1, \dots, q_t) = 1$.

Finally, we need to account for the fact that, for a pattern of size t, each corresponding class has exactly t distinct entries. If, for example, $P = \langle 2, 1, 1\rangle$, then $C = [1, 1, 3, 5]$ is one possible class. However, $[3, 3, 3, 1]$, although it is a valid solution to $\sum_{i=1}^{3} p_i x_i = n$, does not fulfill the requirement that $x_i \ne x_j$ for all $i \ne j$. This is easily seen if we adopt the notation $C = [x_1^{p_1}, \dots, x_t^{p_t}]$. Written in this way,

[3]Note that the number of good partitions of k is equal to the number of partitions of $k-1$, denoted $p(k-1)$. Since $p(k) \sim e^{\pi\sqrt{2k/3}}/4\sqrt{3}k \sim p(k-1)$, almost all patterns are good.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/14/23 to 129.219.247.33 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

$[1,1,3,5] = [1^2 3^1 5^1]$ and $[3,3,3,1] = [3^2 3^1 1^1]$. In the latter case, $x_1 = x_2$, so the class $[3^2 3^1 1^1]$ is not valid. Therefore, we must count the number of such nondistinct solutions. If $x_i = x_k$, then the equation reduces to $(p_i + p_k)x_i + \sum_{j=1, j\neq i, j\neq k}^{t} p_j x_j = n$. We can then apply Schur's theorem to this equation as we did before. For the new system, $Q' = (p_i + p_k)\prod_{j=1, j\neq i, j\neq k}^{t} p_j = Q\frac{p_i+p_j}{p_i p_j}$, so the total number of nondistinct solutions to this equation is $\sim \frac{(n-k)^{t-2}}{(t-2)!Qg^{t-2}} \sum_{i<j} \frac{p_i p_j}{p_i+p_j}$ $\square$

In order to count the classes, we now apply Lemma 8 to each pattern. We want to show that as n gets large, a given bad pattern P has far fewer classes than its good component $\phi(P)$. First, we must determine the size of $\sum_{i<j} \frac{p_i p_j}{p_i+p_j}$ for any pattern P.

For any p_i, we know that $\frac{p_i p_j}{p_i+p_j} = \frac{p_i(c-p_i)}{c}$ for some $c < k$, which is maximized when $p_i = c/2$. If $p_i = c/2$, then $\sum_{j=1}^{t} p_i = \frac{tc}{2} = k$. Thus $c = \frac{2k}{t}$. Since each $p_i = c/2$, we maximize the value of this expression when $p_i = k/t$. Thus, $\sum_{i<j} \frac{p_i p_j}{p_i+p_j} \leq \binom{t}{2}\frac{(k/t)^2}{2k/t} \sim \frac{t^2}{2}\frac{k}{2t} = \frac{kt}{4} \leq k^2/4$.

Using this upper bound, we calculate that the number of classes belonging to a pattern P is

$$c(P) \sim \frac{(n-k)^{t-1}}{(t-1)!Qg^{t-1}} - \frac{(n-k)^{t-2}}{(t-2)!Qg^{t-2}}\frac{k^2}{4} \sim \frac{(n-k)^{t-1}}{(t-1)!Qg^{t-1}}.$$

With this application of Schur's theorem, we can count the number of classes belonging to good patterns compared to the number of classes belonging to bad patterns as follows:

$$\frac{c(P)}{c(\phi(P))} \sim \frac{(n-k)^{t-1}}{(t-1)!g^{t-1}\prod_{j=1}^{t} p_j} \frac{t!\prod_{j=1}^{t-1} p_j (p_t - 1)}{(n-k)^t} \sim \frac{t(p_t-1)}{p_t n} \to 0.$$

That is, the good classes asymptotically outnumber the bad classes.

It still remains to show that the number of forms of a bad class C belonging to a bad pattern P is no greater than the number of forms of a good class C of $\phi(P)$. As we know, the forms of a class are all permutations of the class entries modulo cyclic rotation. Therefore, each class $C \in P$ has $\frac{(k-1)!}{\prod_{j=1}^{t} p_j!}$ forms, while classes $C \in \phi(P)$ have $\frac{k!}{\prod_{j=1}^{t-1} p_j!(p_t-1)}$ forms. That is, the good forms outnumber the bad forms by a factor of kp_t.

Finally, it remains to count the sets. By definition, all good forms have at least one unique element. Thus the good sets have at least one unique difference between two elements. This implies that every good form contains exactly n sets.[4] Therefore, the number of sets per good form is at least the number of sets per bad form.

By counting the patterns, classes, forms, and sets, one can see that almost every set is good. However, it remains to be shown that asymptotically many of these good sets can be arranged into a Ucycle packing. We will use the following lemmas to prove that this is true, and therefore a near-Ucycle packing is always possible.

LEMMA 9. *If $\mathcal{T}_{n,k}$ is restricted to the good classes, then it is a union of cycles, and hence even.*

Proof. For a given class C and transition graph $\mathcal{T}_{n,k}$, we define the graph $\mathcal{T}_{n,k}(C)$ to be the restriction of $\mathcal{T}_{n,k}$ to the edges belonging to the forms of C.

[4]The bad forms do not have a unique element, so this is not always the case for bad sets. Instead, we know only that the number of sets contained in a bad form is a factor of n. (Symmetry can reduce the number of sets; e.g., $\langle 3\rangle$ has forms $(1,4,7)$, $(2,5,8)$, and $(3,6,9)$ when $n = 9$.)

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/14/23 to 129.219.247.33 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

If the class C is good, then $\mathcal{T}_{n,k}(C)$ is a cycle or union of cycles. Each form of C has a unique representative c_k; thus for any permutation F of $\{c_1, \ldots, c_{k-1}\}$, all cyclic permutations of F are also forms of C. Since $\mathcal{T}_{n,k}(C)$ is a union of cycles for each good C, it is clear that $\mathcal{T}_{n,k}$ restricted to the good classes will be a union of Eulerian subgraphs and is therefore even. ∎

COROLLARY 10. *If the restriction of $\mathcal{T}_{n,k}$ to good classes is connected, then it is Eulerian.*

Proof. Since the restriction of $\mathcal{T}_{n,k}$ to good classes is a union of cycles, it is easily seen that if $\mathcal{T}_{n,k}$ is connected, it is surely Eulerian. ∎

By the result in [2], if $\mathcal{T}_{n,k}$ is Eulerian, then an (n,k)-UCS exists. Therefore, if we can show that the restriction of $\mathcal{T}_{n,k}$ to good classes is connected, then it follows that a Ucycle packing exists that includes all good sets. However, Figure 2 demonstrates that this is not always the case. Instead, we prove that an asymptotically large component is connected. Since we have proven that $(1-o(1))\binom{n}{k}$ sets are good, a Ucycle packing that includes all good sets is a near-packing.

3.2. Finding a large component. In order to study the components of the restriction of $\mathcal{T}_{n,k}$ to good classes, we define the *class graph*, denoted $\mathcal{H}_{n,k}$, as the undirected graph whose vertices are all classes of (n,k)-UCS. An edge is drawn between the class representations that differ by only one entry. For example, $[1,2,2;5]$ and $[1,1,2;6]$ are connected in $\mathcal{H}_{10,4}$.

If C_1 and C_2 are connected in $\mathcal{H}_{n,k}$, then this means that $\mathcal{T}_{n,k}(C_1)$ and $\mathcal{T}_{n,k}(C_2)$ will also share vertices. For example, $[1,2,2;5]$ and $[1,1,2;6]$ are connected in $\mathcal{H}_{10,4}$, and correspond to the cycles

$$((12)) \to ((22)) \to ((21)) \to ((12))$$

and

$$((11)) \to ((12)) \to ((21)) \to ((11)).$$

Just because C_1 and C_2 are connected in $\mathcal{T}_{n,k}$, this does not guarantee that the union of $\mathcal{T}_{n,k}(C_1)$ and $\mathcal{T}_{n,k}(C_2)$ will be connected. It could happen, for example, that each C_i has two components that connect, resulting in two components for $C_1 \cup C_2$. However, as proven in [2], if $\mathcal{H}_{n,k}$ is connected, then the union over all classes produces a connected $\mathcal{T}_{n,k}$. We will clarify this with the map κ, defined as follows.

Let $C = [c_1^{p_1}, \ldots, c_{t-1}^{p_{t-1}}; c_t]$ be a class, where the entry c_i appears p_i times and c_t is the largest singleton. To connect the good classes, we define the map $\kappa : c(\Gamma) \to c(\Gamma)$ by $[c_1^{p_1}, \ldots, c_{t-1}^{p_{t-1}}; c_t] \to [1, c_1^{p_1}, \ldots, c_{t-1}^{p_{t-1}-1}; c_t + c_{t-1} - 1]$. We then apply κ again, each time adding another 1.

For example, κ connects the class $[2,2,2;4]$ to the class $[1,1,1;7]$ of $\mathcal{H}_{10,4}$ by the path

$$[2,2,2;4] \to [1,2,2;5] \to [1,1,2;6] \to [1,1,1;7].$$

In this way, we are able to connect the majority of the classes to the class $[1, \ldots, 1; k-t+1]$. However, using κ does not work for every good class. For example, the class $[3,3,3;1]$ is surely good. However, κ maps $[3,3,3;1]$ to $[1,3,3;3]$, that is, to itself. In general, if c_t is the largest singleton of a class C, then $c_t + c_{t-1} - 1$ will be the largest singleton of $\kappa(C)$ only if $c_t > 1$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

3.3. Counting awesome sets. In order to circumvent the problem introduced above, we restrict our attention to the *awesome classes*, the good classes whose largest singleton is greater than 1. Using Schur's theorem, we can show that the number of classes that are not awesome is negligible, as follows:

$$\begin{aligned}
&\left|\left\{(x_1,\dots,x_t) \mid \sum_{j=1}^{t} p_j x_j = n,\ 1 = x_1 = p_1 < p_2 < \cdots < p_t,\ x_j > 1\right\}\right| \\
&\quad = \left|\left\{(x_2,\dots,x_t) \mid \sum_{j=2}^{t} p_j x_j = n-1,\ 1 < p_2 < \cdots < p_t,\ x_j > 1\right\}\right| \\
&\quad = \left|\left\{(y_2,\dots,y_t) \mid \sum_{j=2}^{t} p_j y_j = n-2k-1,\ 1 < p_2 < \cdots < p_t,\ y_j \geq 0\right\}\right| \\
&\quad \sim \frac{(n-2k-1)^{t-2}}{(t-2)!g^{t-2}Q} \\
&\quad \ll n^{t-1}
\end{aligned}$$

for $n > 2k$.

Since the number of nonawesome classes is negligible compared to the number of total classes, we can disregard them and restrict our attention to the awesome classes, which still comprise an asymptotically large proportion of all subsets.

3.4. Proof of Theorem 4. Since almost all classes are awesome, we know that each awesome class is connected to $[1, 1, \dots, 1; n-k+1]$ in $\mathcal{H}_{n,k}$. Since the awesome classes of $\mathcal{H}_{n,k}$ are connected, the restriction of $\mathcal{T}_{n,k}$ to the awesome classes is also connected. If the restriction of $\mathcal{T}_{n,k}$ to awesome classes is connected, then by Lemma 9, we know that all *awesome sets*, the sets belonging to awesome classes, can be connected to form a Ucycle packing. Since, as shown above, the awesome sets represent an asymptotically large proportion of the total sets, it follows that this Ucycle packing is a near-Ucycle packing. $\square$

4. Remarks. Many of the techniques presented here can be extended to other forms of Ucycle approximations. For example, it is possible to include any set in a Ucycle packing by simply inserting the set elements anywhere in the cycle. We could therefore construct a *Ucycle covering* by simply appending all nonawesome sets onto a near-Ucycle created using the techniques we have described. However, this is very inefficient because it increases the length of the cycle by k for each added set instead of the desired 1. In order to find a more elegant method of constructing Ucycle coverings, more complicated analysis is required. As stated earlier, each bad form may produce only a factor of n sets. Therefore, the method used to connect awesome sets into the Ucycle will not work, since traversing an Eulerian transition graph n times is impossible for many bad forms.

One possible method of connecting bad sets in a Ucycle is to consider multiple form classes. Currently, for a form $F = (f_1, \dots, f_k)$, we require $\sum_{i=1}^{k} f_i = n$. However, it could be useful to consider $\sum_{i=1}^{k} f_i = \alpha n$ for $\alpha > 1$. This would allow much more freedom in representing sets, and therefore more ways of connecting sets.

Also, we remark that it is certainly possible to ignore fewer sets than we have in our construction. In the technique we described above, we defined the awesome classes as all those whose largest singleton is greater than 1, and ignored all others. However,

Downloaded 03/14/23 to 129.219.247.33 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

the removal of these other classes may not actually be necessary. By changing the method of connecting the class graph, it should be possible to ignore only a small fraction of the nonawesome classes. In fact, it might be possible to connect the class graph after removing only classes of the form $[1; 2, 2, \ldots, 3, \ldots, 4, \ldots, k]$ whose non-singleton entries differ by exactly 1, and classes of the form $[1; k, k, \ldots, k]$. However, such considerations further complicate the analysis required without much gain in result. It is reasonably interesting and certainly worthwhile to consider the fewest number of sets necessary to ignore (hopefully zero!) in a near-Ucycle packing, but our goal here was just to show that the number was small.

Finally, due to our proof that near-Ucycles exist, we believe that we deserve asymptotically much of the prize money, or $\$[1 - o(1)](250.04)$. Since we do not know the speed of the $o(1)$ term, we have made a conservative estimate of \$249.99.

REFERENCES

[1] F. R. K. Chung, P. Diaconis, and R. I. Graham, *Universal cycles for combinatorial structures*, Discrete Math., 110 (1992), pp. 43–59.

[2] G. Hurlbert, *On universal cycles for k-subsets of an n-set*, SIAM J. Discrete Math., 7 (1994), pp. 598–604.

[3] B. Jackson, *Universal cycles of k-subsets and k-permutations*, Discrete Math., 117 (1993), pp. 141–150.

[4] B. Jackson, private communication, 2004.

[5] D. E. Knuth, *The Art of Computer Programming, Vol.* 4*, Fasc.* 3*. Generating All Combinations and Permutations*, Addison–Wesley, Upper Saddle River, NJ, 2005.

[6] F. Ruskey, *Combinatorial Generation*, in preparation.

[7] C. Savage, *A survey of combinatorial Gray codes*, SIAM Rev., 39 (1997), pp. 605–629.

[8] B. Stevens, P. Buskell, P. Ecimovic, C. Ivanescu, A. Malik, A. Savu, T. Vassilev, H. Verrall, B. Yang, and Z. Zhao, *Solution of an outstanding conjecture: The nonexistence of universal cycles with $k = n - 2$*, Discrete Math., 258 (2002), pp. 193–204.

[9] H. S. Wilf, *generatingfunctionology*, 2nd ed., Academic Press, Boston, 1994.

Downloaded 03/14/23 to 129.219.247.33 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.