

Generalized de Bruijn Cycles

Joshua N. Cooper¹ and Ronald L. Graham²

¹Department of Mathematics, Courant Institute of Mathematical Sciences, New York University
NY 10012, USA
cooper@cims.nyu.edu

²Department of Computer Science and Engineering, University of California at San Diego, CA
92093, USA
graham@ucsd.edu

Received February 5, 2004

AMS Subject Classification: 94A55, 05C70

Abstract. For a set of integers \mathcal{I} , we define a q -ary \mathcal{I} -cycle to be an assignment of the symbols 1 through q to the integers modulo q^n so that every word appears on some translate of \mathcal{I} . This definition generalizes that of de Bruijn cycles, and opens up a multitude of questions. We address the existence of such cycles, discuss “reduced” cycles (ones in which the all-zeroes string need not appear), and provide general bounds on the shortest sequence which contains all words on some translate of \mathcal{I} . We also prove a variant on recent results concerning decompositions of complete graphs into cycles and employ it to resolve the case of $|\mathcal{I}| = 2$ completely.

Keywords: de Bruijn cycle, graph decomposition, probabilistic method

1. Introduction

A *de Bruijn cycle* of order n is a q -ary sequence $(S(0), \dots, S(q^n - 1))$ so that every q -ary n -word appears in a “window” $(S(j), \dots, S(j + n - 1))$ for some j (indices taken modulo q^n). A *reduced* de Bruijn cycle is a string of length $q^n - 1$ which achieves every n -word in some window, except for the word 0^n . In this paper, we are concerned with such objects when the notion of “window” is generalized.

Given a sequence $\mathcal{I} = \{i_j\}_{j=1}^n \subset \mathbb{Z}_{q^n}$, we say that the map $\chi: \mathbb{Z}_{q^n} \rightarrow [q]$ (respectively, $\chi: \mathbb{Z}_{q^n-1} \rightarrow [q]$) is an \mathcal{I} -cycle (\mathcal{I}^* -cycle) if, for every word $W \in [q]^n$ (resp., every word $W \in [q]^n \setminus \{0^n\}$), there exists a $t \in \mathbb{Z}_{q^n}$ (respectively, $t \in \mathbb{Z}_{q^n-1}$) so that $\chi(i_j + t) = W(j)$. If such a sequence exists for \mathcal{I} , we say that \mathcal{I} is q -valid (resp., q^* -valid). We will often abuse notation by writing $\chi(\mathcal{I})$ for the n -word $(\chi(i_1), \dots, \chi(i_n))$. Furthermore, we *also* use “cycle” to refer to sequences of edges in a directed graph each of whose tail is the head of the previous one, and which return to their starting point. It should always be clear from context which of these definitions we intend — though, often, the notions will coincide!

It is classical that an \mathcal{I} -cycle and an \mathcal{I}^* -cycle exist for $\mathcal{I} = \{1, \dots, n\}$. In Section 2, we address the validity of other sets \mathcal{I} . As it turns out, the question is rather difficult in general, and we solve the problem completely only for sets of cardinality 2. We present some general constructions and a number of computational results, and we discuss a graph-theoretic question whose solution is equivalent to the case of \mathcal{I} being an arithmetic progression. In the next section, we discuss the existence of reduced de Bruijn cycles, and have a greater degree of success in characterizing the q^* -valid sets. Then, in Section 4, the issue of “approximate” cycles is discussed, and we present a nearly optimal bound on their length. The following section contains a proof of a graph-theoretic decomposition result that is used in Section 2 and which solves a variant of a family of problems that has appeared recently in the literature. We finish with a number of open questions and suggestions for future investigation.

2. Unreduced Cycles

It is easy to see that, to determine the two-element q -valid sets, we need only examine the sets $\{0, d\}$ with $d|q^2$. Indeed, for any $k \in \mathbb{Z}_q^\times$, if there exists an \mathcal{I} -cycle χ for $\{0, d\}$, then $\chi': s \rightarrow \chi(k^{-1}s)$ is an \mathcal{I} -cycle for $\{0, kd\}$. In addition, it is clear that the validity of \mathcal{I} is equivalent to the validity of $\mathcal{I} + b$ for any $b \in \mathbb{Z}_q$. The same arguments apply to sets \mathcal{I} whose elements are in arithmetic progression: we need only examine the cases when the difference d divides q^n .

Suppose, then, that $d|q^n$, and consider D_q^n , the n^{th} q -ary *de Bruijn digraph*, i.e., the digraph whose vertices are the q -ary n -strings and which has an edge from x to y if the last $n-1$ symbols of x are the same as the first $n-1$ symbols of y . (Note that some edges have loops attached.) Then the set $AP(n, d) = \{0, d, 2d, \dots, (n-1)d\}$ is q -valid iff there is a partition of the edges of D_q^{n-1} into d cycles each of length q^n/d , because we may write $\chi(j) = C_a(b)$, where C_a is the a^{th} such cycle and $j = at + b$ with $t = q^n/d$.

Using D_q^1 , which is simply a complete directed graph with loops on q vertices, we may state a condition equivalent to the q -validity of $\{0, d\}$: that D_q^1 is fully decomposable into cycles of length q^2/d . Theorem 5.5, which appears in Section 5, says that this is possible precisely when $q^2/d > 2$. Therefore, we have

Theorem 2.1. *There exists an \mathcal{I} -cycle for $\mathcal{I} = \{0, d\}$ if and only if $q^2/d \neq 2$.*

The situation for sets with $|\mathcal{I}| > 2$ appears significantly more complicated, even for arithmetic progressions. However, the above invalidity result for $d = q^2/2$ has an immediate analogue for larger n :

Proposition 2.2. *For any r, q , with $r|q$, the set $AP(r, q^r/r)$ is q -invalid.*

Proof. Suppose an \mathcal{I} -cycle χ existed. Then, we may assume without loss of generality that $(\chi(0), \dots, \chi((r-1)q^n/r))$ is the all-zeroes vector. But, then $(\chi(q^r/r), \chi(2q^r/r), \dots, \chi(0))$ is also, a contradiction. ■

On the other hand, we can construct a large family of $AP(n, q)$ -cycles. Form the *quotient graph* $G(n)$ from D_q^n by identifying two vertices x and x' of D_q^n if $x - x' = k$ for some $k \in \mathbb{Z}_q$, i.e., $x_i - x'_i = k$ for $1 \leq i \leq n$.

Fact 1. $G(n)$ is isomorphic to D_q^{n-1} .

Proof. For $x = (x_1, x_2, \dots, x_n)$, consider the map

$$\lambda: x \mapsto (x_1 - x_2, x_2 - x_3, \dots, x_{n-1} - x_n).$$

It is easy to check that λ is well-defined on $G(n)$, invertible and preserves directed edges (i.e., (x, y) is an edge in D_q^n if and only if $(\lambda(x), \lambda(y))$ is an edge in $G(n)$). Note that the inverse map λ^{-1} does not necessarily preserve cycles, though. However, it is not hard to show the following. Suppose that $(q_0, q_1, \dots, q_{r-1})$ is a cycle in $G(n) \cong D_q^{n-1}$, i.e.,

$$((q_{i+1}, q_{i+2}, \dots, q_{i+n-1}), (q_{i+2}, q_{i+3}, \dots, q_{i+n}))$$

is an edge for all i , where the indices are reduced modulo r . Then this cycle “lifts” under λ^{-1} to a cycle in D_q^n if and only if $\sum_{i=0}^{r-1} q_i = 0 \pmod{q}$.

Observe that any Eulerian cycle C in D_q^{n-1} satisfies this sum condition. Hence, it lifts to a cycle C^+ in D_q^n going through exactly one of the q points in each equivalence class. In fact, we can form q disjoint cycles $\{C_j^+\}_{j=1}^q$ from this cycle C^+ by translating each point in it by some fixed constant $q \in \mathbb{Z}_q$.

Finally, we can form a cyclic sequence S in D_q^n containing all of its vertices by “splicing together” these q cycles C_j^+ in the obvious way. Since C was in fact a de Bruijn cycle for $(n-1)$ -tuples, then it easily checked that S is an \mathcal{I} -cycle with $\mathcal{I} = \{0, q, 2q, \dots, (n-1)q\}$. ■

As an example, consider the cycle 001122021 for $n = 2, q = 3$. We can lift this to 100021200 for $n = 3, q = 3$, form the two translates 211101011 and 022210122 and splice them together to get 021210210210102021102210210, which is a $\{0, 3, 6\}$ -cycle.

In fact, since there are many ways of choosing the first de Bruijn cycle ($[(q-1)!]^{q^{n-1}} \cdot q^{q^{n-2}-n+1}$, to be precise), and many ways of splicing them together ($q!$), there are quite a few ways of producing such cycles. Unfortunately, it is not possible to iterate this construction, since the cycles C_q^+ do not themselves have the zero-sum property needed to lift them again.

Proposition 2.3. $AP(n, q)$ is q -valid for any n, q .

Suppose that $\{0, d, 2d\}$ is q -valid, where $8|k = q^3/d$. Then there exists a decomposition of the edges of D_q^2 into cycles of length k . Write \mathcal{E} for the set of even symbols in $[2q]$ and \mathcal{O} for the set of odd symbols. We may think of D_{2q}^n as being composed of four parts: $U_1 = \mathcal{E} \times \mathcal{E}$, $U_2 = \mathcal{E} \times \mathcal{O}$, $U_3 = \mathcal{O} \times \mathcal{O}$, and $U_4 = \mathcal{O} \times \mathcal{E}$. U_2 and U_4 contain no edges; U_1 and U_3 are copies of D_q^2 . We may therefore decompose U_1 and U_3 into k -cycles. The remaining edges may be decomposed into 4-cycles and 8-cycles as follows. For each pair $(a, b), (b, c)$ with $a, b \in \mathcal{E}$ and $c \in \mathcal{O}$, define a cycle $C(a, b, c) = ((a, b), (b, c), (c, a+b+c), (a+b+c, a))$, with addition modulo $2q$. The resulting 4-cycles partition all edges which do not belong to $U_1 \times U_1, U_3 \times U_3, U_2 \times U_4$, or $U_4 \times U_2$. The edges in these final two classes come in pairs $\{(b, c), (c, b)\}$. We may attach $\{(b, c), (c, b)\}$ and $\{(2b+c+2, b), (b, 2b+c+2)\}$ to the cycle $C(b+2, b, c)$

for each b even and $c = 1 \pmod{4}$, thus turning it into an 8-cycle. Doing so accounts for all the remaining edges exactly once.

The result is a set of 8-cycles and 4-cycles. We may partition them into classes so that each class has exactly k edges, and “join” each class at U_1 into a cycle of length k . The resulting decomposition of D_{2q}^n gives rise to a $\{0, 8d, 16d\}$ -cycle. Hence, we have the following.

Proposition 2.4. *If $AP(3, d)$ is q -valid, where $8|q^3/d$, then $AP(3, 8d)$ is $2q$ -valid.*

Corollary 2.5. *$AP(3, 8^k)$ is $r2^{k+1}$ -valid for all $k \geq 0$ and $r \geq 1$.*

Proof. Begin with the $2r$ -valid $\{0, 1, 2\}$ and iterate the above proposition. ■

To illustrate the complicated nature of the $d > 2$ case, we offer the following computational observations. By “affine equivalence”, we mean a map $\sigma: s \rightarrow ks + b$ for some $k \in \mathbb{Z}_{q^n}^\times$ and $b \in \mathbb{Z}_{q^n}$. Clearly, the partition of index sets into valid and invalid is refined by the partition into affine equivalence classes.

1. For $(q, n) = (3, 3)$, the only invalid index sets are $\{k, k+9, k+18\}$ for $k = 0, \dots, 8$.
2. The only 2-invalid 3-set (up to affine equivalence) is $\{0, 1, 3\}$. For $(q, n) = (2, 4)$, the only valid \mathcal{I} 's (up to affine equivalence) are the nine sets $\{0, 1, 2, 3\}$, $\{0, 1, 2, 6\}$, $\{0, 1, 2, 7\}$, $\{0, 1, 3, 4\}$, $\{0, 1, 3, 7\}$, $\{0, 1, 3, 8\}$, $\{0, 1, 3, 9\}$, $\{0, 1, 3, 14\}$, and, of course, $\{0, 2, 4, 6\}$.
3. For $(q, n) = (2, 5)$, the following list contains one representative of each equivalence class of invalid index sets, in lexicographic order:

0, 1, 2, 3, 12	0, 1, 2, 4, 7	0, 1, 2, 4, 9	0, 1, 2, 4, 12	0, 1, 2, 4, 23
0, 1, 2, 4, 24	0, 1, 2, 4, 25	0, 1, 2, 4, 26	0, 1, 2, 4, 27	0, 1, 2, 5, 7
0, 1, 2, 5, 8	0, 1, 2, 5, 9	0, 1, 2, 5, 13	0, 1, 2, 5, 14	0, 1, 2, 5, 15
0, 1, 2, 5, 16	0, 1, 2, 5, 19	0, 1, 2, 5, 20	0, 1, 2, 5, 21	0, 1, 2, 5, 22
0, 1, 2, 5, 24	0, 1, 2, 5, 25	0, 1, 2, 5, 26	0, 1, 2, 6, 9	0, 1, 2, 6, 11
0, 1, 2, 6, 13	0, 1, 2, 6, 14	0, 1, 2, 6, 15	0, 1, 2, 6, 16	0, 1, 2, 6, 17
0, 1, 2, 6, 19	0, 1, 2, 6, 21	0, 1, 2, 6, 23	0, 1, 2, 6, 25	0, 1, 2, 6, 26
0, 1, 2, 7, 11	0, 1, 2, 7, 14	0, 1, 2, 7, 15	0, 1, 2, 7, 19	0, 1, 2, 7, 22
0, 1, 2, 7, 23	0, 1, 2, 7, 24	0, 1, 2, 8, 12	0, 1, 2, 8, 13	0, 1, 2, 8, 16
0, 1, 2, 8, 17	0, 1, 2, 8, 18	0, 1, 2, 8, 19	0, 1, 2, 8, 23	0, 1, 2, 8, 24
0, 1, 2, 8, 25	0, 1, 2, 9, 13	0, 1, 2, 9, 14	0, 1, 2, 9, 15	0, 1, 2, 9, 16
0, 1, 2, 9, 17	0, 1, 2, 9, 19	0, 1, 2, 9, 20	0, 1, 2, 9, 21	0, 1, 2, 9, 22
0, 1, 2, 9, 25	0, 1, 2, 10, 14	0, 1, 2, 10, 15	0, 1, 2, 10, 16	0, 1, 2, 10, 17
0, 1, 2, 10, 18	0, 1, 2, 10, 20	0, 1, 2, 11, 13	0, 1, 2, 11, 14	0, 1, 2, 11, 15
0, 1, 2, 11, 16	0, 1, 2, 11, 19	0, 1, 2, 12, 15	0, 1, 2, 12, 17	0, 1, 2, 12, 19
0, 1, 2, 13, 16	0, 1, 2, 13, 18	0, 1, 2, 13, 20	0, 1, 2, 14, 17	0, 1, 2, 15, 18
0, 1, 2, 16, 18	0, 1, 3, 4, 9	0, 1, 3, 4, 11	0, 1, 3, 4, 12	0, 1, 3, 4, 15
0, 1, 3, 4, 16	0, 1, 3, 5, 9	0, 1, 3, 5, 11	0, 1, 3, 5, 12	0, 1, 3, 5, 13
0, 1, 3, 5, 15	0, 1, 3, 5, 17	0, 1, 3, 5, 21	0, 1, 3, 5, 22	0, 1, 3, 5, 24
0, 1, 3, 5, 25	0, 1, 3, 5, 26	0, 1, 3, 7, 9	0, 1, 3, 7, 11	0, 1, 3, 7, 12
0, 1, 3, 7, 15	0, 1, 3, 7, 16	0, 1, 3, 7, 17	0, 1, 3, 7, 19	0, 1, 3, 7, 23

0, 1, 3, 7, 24	0, 1, 3, 7, 27	0, 1, 3, 7, 30	0, 1, 3, 8, 10	0, 1, 3, 8, 12
0, 1, 3, 8, 14	0, 1, 3, 8, 16	0, 1, 3, 8, 17	0, 1, 3, 8, 19	0, 1, 3, 8, 20
0, 1, 3, 8, 21	0, 1, 3, 8, 23	0, 1, 3, 8, 24	0, 1, 3, 9, 13	0, 1, 3, 9, 16
0, 1, 3, 9, 17	0, 1, 3, 9, 20	0, 1, 3, 9, 25	0, 1, 3, 9, 26	0, 1, 3, 9, 28
0, 1, 3, 9, 30	0, 1, 3, 10, 12	0, 1, 3, 10, 13	0, 1, 3, 10, 14	0, 1, 3, 10, 15
0, 1, 3, 10, 16	0, 1, 3, 10, 20	0, 1, 3, 10, 23	0, 1, 3, 10, 30	0, 1, 3, 12, 13
0, 1, 3, 12, 16	0, 1, 3, 12, 24	0, 1, 3, 12, 25	0, 1, 3, 12, 27	0, 1, 3, 12, 28
0, 1, 3, 13, 15	0, 1, 3, 13, 21	0, 1, 3, 13, 22	0, 1, 3, 13, 25	0, 1, 3, 13, 27
0, 1, 3, 13, 28	0, 1, 3, 14, 15	0, 1, 3, 15, 16	0, 1, 3, 15, 20	0, 1, 3, 15, 21
0, 1, 3, 15, 22	0, 1, 3, 15, 23	0, 1, 3, 15, 25	0, 1, 3, 15, 27	0, 1, 3, 15, 28
0, 1, 3, 16, 17	0, 1, 3, 16, 19	0, 1, 3, 16, 21	0, 1, 3, 16, 23	0, 1, 3, 16, 25
0, 1, 3, 16, 27	0, 1, 3, 17, 19	0, 1, 3, 17, 21	0, 1, 3, 17, 23	0, 1, 3, 17, 25
0, 1, 3, 17, 27	0, 1, 3, 17, 28	0, 1, 3, 19, 27	0, 1, 3, 20, 24	0, 1, 3, 21, 24
0, 1, 3, 21, 25	0, 1, 3, 22, 24	0, 1, 3, 22, 25	0, 1, 3, 23, 24	0, 1, 3, 23, 25
0, 1, 3, 23, 27	0, 1, 3, 24, 28	0, 1, 3, 25, 27	0, 1, 3, 27, 28	0, 1, 4, 5, 13
0, 1, 4, 6, 12	0, 1, 4, 6, 14	0, 1, 4, 6, 17	0, 1, 4, 6, 18	0, 1, 4, 6, 20
0, 1, 4, 8, 23	0, 1, 4, 9, 15	0, 1, 4, 9, 16	0, 1, 4, 9, 17	0, 1, 4, 9, 20
0, 1, 4, 12, 14	0, 1, 4, 12, 18	0, 1, 4, 13, 14	0, 1, 4, 14, 17	0, 1, 4, 14, 28
0, 1, 4, 14, 29	0, 1, 4, 15, 16	0, 1, 4, 15, 20	0, 1, 4, 15, 23	0, 1, 4, 15, 28
0, 1, 4, 16, 26	0, 1, 4, 17, 26	0, 1, 4, 18, 26	0, 1, 4, 26, 28	0, 1, 5, 7, 16
0, 1, 5, 8, 16	0, 1, 6, 8, 17	0, 1, 7, 8, 17	0, 1, 7, 9, 15	0, 1, 7, 9, 16
0, 1, 7, 9, 17	0, 1, 7, 15, 16	0, 1, 8, 16, 17	0, 1, 8, 16, 24	0, 2, 4, 8, 14
0, 2, 4, 10, 14	0, 2, 4, 10, 18	0, 2, 4, 10, 24	0, 2, 4, 10, 26	0, 2, 4, 16, 20
0, 2, 6, 18, 22	0, 2, 8, 16, 18	0, 2, 8, 16, 24	0, 4, 8, 16, 24	

3. Reduced Cycles

Although the definition of q^* -validity certainly makes sense when q is not a prime power, we restrict our attention to that case in this section. Therefore, consider $q \geq 2$ a fixed prime power, and take our alphabet to be \mathbb{F}_q . Let α be a generator of the multiplicative group of the finite field \mathbb{F}_{q^n} . Denote by \mathcal{E} the elementary basis for \mathbb{F}_q^n over \mathbb{F}_q . Given a basis $\mathcal{B} = \{b_1, \dots, b_n\}$ of \mathbb{F}_{q^n} over \mathbb{F}_q and an element $\gamma \in \mathbb{F}_{q^n}$, write $f_{\mathcal{B}}(\gamma)$ for the element of \mathbb{F}_q^n whose j^{th} coordinate is the coefficient of b_j in the \mathcal{B} -representation of γ . Then, given a nonzero vector $\mathbf{v} \in \mathbb{F}_q^n$, define $\Lambda(\alpha, \mathcal{B}, \mathbf{v})$ to be the string whose j^{th} coordinate (i.e., $\Lambda_j(\alpha, \mathcal{B}, \mathbf{v})$, $0 \leq j \leq q^n - 2$) is $\mathbf{v}^T f_{\mathcal{B}}(\alpha^j)$.

It is well known that, when $\mathcal{B} = \{\alpha^j : 0 \leq j \leq n - 1\}$ and \mathbf{v} has only one nonzero coordinate, $\Lambda(\alpha, \mathcal{B}, \mathbf{v})$ is a reduced de Bruijn cycle of order n (e.g., [6]). We generalize this result as follows.

Proposition 3.1. *Let $\mathcal{I} = \{i_j\}_{j=1}^n$ be a sequence of distinct integers. Fix a basis \mathcal{B} of \mathbb{F}_q^n over \mathbb{F}_q , a generator $\alpha \in \mathbb{F}_{q^n}^\times$, and a vector $\mathbf{v} \in \mathbb{F}_q^n$, and write $\Phi(t)$ for the vector*

$$(\Lambda_{i_1+t}(\alpha, \mathcal{B}, \mathbf{v}), \dots, \Lambda_{i_n+t}(\alpha, \mathcal{B}, \mathbf{v}))^T \in \mathbb{F}_q^n,$$

with indices taken modulo $q^n - 2$. If the minimal polynomial of α is not a divisor of $\sum_{j=1}^n c_j x^{i_j}$ for any nonzero (c_1, \dots, c_n) , then the map Ψ which sends 0 to 0 and α^t to $\Phi(t)$ is an isomorphism from the additive group of \mathbb{F}_{q^n} to \mathbb{F}_q^n .

Proof. First, we show that Ψ is linear. Write e_j for the elementary n -vector whose coordinates are all zero except for a 1 in the j^{th} coordinate. We denote by $M_{\gamma, \mathcal{B}}$ the matrix representing multiplication by $\gamma \in \mathbb{F}_{q^n}$ in the \mathcal{B} basis. It is easy to see that

$$\Lambda_k(\alpha, \mathcal{B}, \mathbf{v}) = \mathbf{v}^\top f_{\mathcal{B}}(\alpha^k),$$

and therefore that

$$\Psi(\gamma) = \sum_{j=1}^n e_j \mathbf{v}^\top f_{\mathcal{B}}(\alpha^{ij} \gamma) = \sum_{j=1}^n e_j \mathbf{v}^\top M_{\alpha, \mathcal{B}}^{ij} f_{\mathcal{B}}(\gamma), \quad (3.1)$$

which is obviously linear.

Now, suppose that $\Psi(\gamma) = 0$ and $\gamma = \alpha^t$. If we denote by S the subspace of \mathbb{F}_q^n orthogonal to \mathbf{v} , then we have $\alpha^{ij+t} \in f_{\mathcal{B}}^{-1}(S)$ for each j . However, $f_{\mathcal{B}}$ is linear and has a trivial kernel, so all the α^{ij+t} lie in a subspace of \mathbb{F}_q^n of dimension $n-1$ and are therefore linearly dependent. Since $M_{\alpha, \mathcal{B}}$ is nonsingular, this implies that $\{\alpha^{ij}\}_{j=1}^n$ is a dependent set. But then we have

$$\sum_{j=1}^n c_j \alpha^{ij} = 0$$

for some nonzero (c_1, \dots, c_n) , a contradiction. \blacksquare

The map $\gamma \mapsto M_{\gamma, \mathcal{B}}$ is actually an isomorphism of fields. The image is a set of matrices which form a field, i.e., a *matrix field*. These objects have been studied extensively and thoroughly characterized when the matrices take their entries from a finite field (see [4]).

Corollary 3.2. *If the minimal polynomial of α , a multiplicative generator of $\mathbb{F}_{q^n}^\times$, is not a divisor of $\sum_{j=1}^n c_j x^{ij}$ for any nonzero (c_1, \dots, c_n) , then $\Lambda(\alpha, \mathcal{B}, \mathbf{v})$ is an \mathcal{I}^* -cycle.*

Proof. By the above argument, $\Lambda(\alpha, \mathcal{B}, \mathbf{v})$ contains all nonzero n -strings in shifted copies of the index set \mathcal{I} . \blacksquare

We require another definition.

Definition 3.3. *The index set $\mathcal{I} = \{i_j\}_{j=1}^n$ is called exceptional for q if, for every primitive polynomial $g \in \mathbb{F}_q[x]$ of degree n , there exists a nonzero vector $(c_1, \dots, c_n) \in \mathbb{F}_q^n$ so that g divides*

$$\sum_{j=1}^n c_j x^{i_j}. \quad (3.2)$$

Equivalently, if for every α a generator of $\mathbb{F}_{q^n}^\times$, the set $\{\alpha^{i_j}\}$ is linearly dependent over \mathbb{F}_q , then \mathcal{I} is exceptional for q . An index set which is not exceptional is called ordinary for q .

Note that the exponents in (3.2) can be thought of as belonging to \mathbb{Z}_{q^n-1} .

Proposition 3.4. *\mathcal{I} is q^* -valid whenever \mathcal{I} is ordinary for q .*

Proof. If $\mathcal{I} = \{i_j\}_{j=1}^n$ is not exceptional, then there exists a primitive polynomial $g \in \mathbb{F}_q[x]$ of degree n so that, for all nonzero (c_1, \dots, c_n) , g is not a divisor of $\sum_{j=1}^n c_j x^{i_j}$. Since x is not a root of $\sum_{j=1}^n c_j x^{i_j}$ in $\mathbb{F}_q[x]/g$, but it is a multiplicative generator of this field, $\Lambda(x, \mathcal{B}, \mathbf{v})$ is an \mathcal{I}^* -cycle for any \mathcal{B} and \mathbf{v} . ■

Which index sets are ordinary? We argue that if $(q^n - 1, d) = 1$, then $(a, a + d, a + 2d, \dots, a + (n-1)d)$ is ordinary. It is clear that, if \mathcal{I} is ordinary, all of its translates are as well. We may therefore assume that $a = 0$. Then some irreducible polynomial g of degree n divides $\sum_{j=0}^{n-1} c_j x^{jd}$ for each $(c_1, \dots, c_n) \neq 0$. Let α be a root of g , so α^d is a root of $\sum_{j=0}^{n-1} c_j x^j$. That this polynomial has degree less than n contradicts the fact that α^d is a generator of $\mathbb{F}_{q^n}^\times$.

It is trivial that \mathcal{I} is ordinary if it is a singleton. If \mathcal{I} has two elements, then it is easy to see that $\mathcal{I} = \{i, j\}$ is ordinary if $(q+1) \nmid (i-j)$, since then $\alpha^{i-j} \notin \mathbb{F}_q$. Indeed, the copy of \mathbb{F}_q^\times lying inside of $\mathbb{F}_{q^2}^\times$ is the set $\{\alpha^{k(q+1)}\}_{k=1}^{q-1}$ for any generator α . Conversely, if $q+1 \mid i-j$, then, for every generator α , we have $\alpha^i = c\alpha^j$ for some $c \in \mathbb{F}_q$. Therefore, a two-element set is ordinary for q if and only if the difference of the elements is not a multiple of $q+1$.

For a prime p and a positive integer n , define the *Jacobi logarithm* as follows: for a generator α of $\mathbb{F}_{p^n}^\times$, define $L_\alpha: \mathbb{Z}_{p^n-1} \setminus \{s\} \mapsto \mathbb{Z}_{p^n-1} \setminus \{0\}$ by $1 + \alpha^t = \alpha^{L_\alpha(t)}$, where $s = (p^n - 1)/2$ if $p > 2$ and $s = 0$ otherwise.

Proposition 3.5. *A three-element set $\{i, j, k\}$ is exceptional if and only if either:*

1. $Q \mid (j-i)$,
2. $Q \mid (k-j)$,
3. $Q \mid (k-i)$,
4. *or, for all m with $(m, q^3 - 1) = 1$, there exists an a so that*

$$L_\alpha(aQ + m(j-i)) \equiv m(k-i) \pmod{Q},$$

where $Q = q^2 + q + 1$, and α is any (fixed) generator of $\mathbb{F}_{q^3}^\times$.

Proof. There are only two ways that α^i , α^j , and α^k can be linearly dependent. Either one of them is an \mathbb{F}_q -multiple of another, or, for some triple $\{c_1, c_2, c_3\}$, with $c_i \neq 0$ for all i ,

$$c_1 \alpha^i + c_2 \alpha^j + c_3 \alpha^k = 0. \quad (3.3)$$

The former case is precisely the divisibility conditions stated above. To see that the latter situation is equivalent to condition 4, we may rewrite (3.3) without loss of generality as

$$c_4 \alpha^{i-k} + c_5 \alpha^{j-k} = 1 \quad (3.4)$$

with $c_4, c_5 \in \mathbb{F}_q$. Suppose $\{i, j, k\}$ falls into this case, i.e., (3.4) has a solution in c_4 and c_5 . Since we have assumed that neither c_4 nor c_5 is zero, we may express each of them in terms of α : respectively, α^{sQ} and α^{tQ} , for some s, t integers. Then we may rewrite (3.4) as

$$\alpha^{sQ+i-k} (1 + \alpha^{Q(t-s)+j-i}) = \alpha^{sQ+i-k+L_\alpha(Q(t-s)+j-i)} = \alpha^0,$$

which is to say, that $sQ + i - k + L_\alpha(Q(t - s) + j - i) \equiv 0 \pmod{q^3 - 1}$. (Note that the fact that L_α is not defined on all of \mathbb{Z}_{q^3-1} is not problematic, since the left hand side of (3.4) cannot be zero.) This equation has a solution in s and t iff there exists an a so that $L_\alpha(aQ + j - i) \equiv k - i \pmod{Q}$.

If we choose any other generator β of $\mathbb{F}_{q^3}^\times$, there is some m such that $(m, q^3 - 1) = 1$ and $\alpha^m = \beta$. Then $L_\beta(a) = b$ iff $L_\alpha(am) = bm$, so $\{i, j, k\}$ is exceptional iff $Q|(j - i)$, $Q|(k - j)$, $Q|(k - i)$, or, for some m such that $(m, q^3 - 1) = 1$, there exists an a so that

$$L_\alpha(aQ + m(j - i)) \equiv m(k - i) \pmod{Q}. \quad \blacksquare$$

4. Approximate Cycles

Since the question of whether an \mathcal{I} -cycle exists appears difficult in general, we may ask instead whether it is possible to find an ‘‘approximate’’ \mathcal{I} -cycle. This question comes in two forms for an index set $\mathcal{I} = \{i_j\}_{j=1}^n \subset \mathbb{Z}$:

1. What is the least N for which a $\chi: \mathbb{Z}_N \rightarrow [q]$ exists so that, for every word W , there exists an m with $W = \chi(\mathcal{I} + m)$?
2. What is the least N for which there exists a χ so that all but $o(q^N)$ words appear as $\chi(\mathcal{I} + m)$ for some m ?

Call the former object a *Type I approximate cycle* and the latter a *Type II approximate cycle*. Then we can show:

Theorem 4.1. *Let $F(n)$ be any function so that $q^{-n}F(n) \rightarrow \infty$ as $n \rightarrow \infty$. Then there exists a q -ary Type II approximate \mathcal{I} -cycle of length $F(n)$ when $|\mathcal{I}| = n$.*

We need a result of Janson to proceed. The following appears in [7]. First, some notation. Let I be an index set for a set of events $\{B_i\}_{i \in I}$. Define a graph \sim on I with the following property: Let J_1 and J_2 be two disjoint subsets of I such that there is no $i_1 \in J_1$ and $i_2 \in J_2$ with $i_1 \sim i_2$. Now, let A^1 be any Boolean function of the events $\{B_i: i \in J_1\}$ and let A^2 be any Boolean function of the events $\{B_i: i \in J_2\}$. Then A^1 and A^2 are independent.

Let $\mu = \sum_{i=1}^m \mathbf{P}(B_i)$, $\Delta = \sum_{i \sim j} \mathbf{P}(B_i \wedge B_j)$, and $\delta = \max_i \sum_{j \sim i} \mathbf{P}(B_j)$. Then the following holds.

Lemma 4.2. *With the above notation,*

$$\mathbf{P}(\wedge_{i=1}^m \overline{B}_i) \leq \exp\left(-\min\left(\frac{\mu^2}{8\Delta}, \frac{\mu}{2}, \frac{\mu}{6\delta}\right)\right).$$

Proof. [Proof of Theorem 4.1.] Fix an integer m . We wish to show that (for a suitable choice of m) the expected number of q -ary n -words which do not appear as $\chi(\mathcal{I} + j)$ for any $j \in \mathbb{Z}_m$ for a random function $\chi: \mathbb{Z}_m \rightarrow q$ is $o(q^n)$. This quantity is q^n times the probability that a single word — say, 0^n — does not appear anywhere. So we need only show that this probability is $o(1)$. Let B_j be the event that $\chi(\mathcal{I} + j) = W$. Then define a graph on the B_j as follows: $B_j \sim B_k$ iff $(\mathcal{I} + j) \cap (\mathcal{I} + k) \neq \emptyset$. Note that $\deg(B_j) \leq n^2$ for all j , and $\mathbf{P}(B_i) = q^{-n}$. Therefore $\mu = mq^{-n}$,

$$\Delta \leq q^n n^2 q^{-n} = n^2,$$

and $\delta \leq n^2 q^{-n}$. Plugging into the lemma, we find

$$\mathbf{P}(\wedge_{i=1}^m \overline{B}_i) \leq \exp\left(-\min\left(\frac{m^2}{8n^2 q^n}, \frac{m}{2q^n}, \frac{m}{6n^2}\right)\right).$$

If we let $m = F(n)$, then $\min(m^2/8n^2 q^n, m/2q^n, m/6n^2) \rightarrow \infty$, completing the proof. ■

Now, we address the problem of constructing Type I approximate \mathcal{I} -cycles. First, for a set of reals $X = \{x_i\}_{i=1}^n$, define

$$\mu(X) = \max_{\alpha \in \mathbb{R}} \min_{j \neq k} \|\alpha(x_j - x_k)\|,$$

where $\|x\|$ denotes the distance to the closest integer. It is proven in [8] that, for any set X of cardinality n , $\mu(X) \geq n^{-2}$. Then we have the following.

Lemma 4.3. *For any $\mathcal{I} = \{i_j\}_{j=1}^n \subset \mathbb{Z}$ and collection of q -ary n -words W_1, \dots, W_S , there exists an integer $p = n^2 S(1 + o(1))$ and a map $\chi: \mathbb{Z}_p \rightarrow [q]$ so that every W_j appears as $\chi(\mathcal{I} + t)$ for some $t \in \mathbb{Z}_p$.*

Proof. Let p be the smallest prime greater than $n^2(S + 3)$, and choose α achieving the bound $\mu(X) \geq n^{-2}$. Then there exists a $k \in \mathbb{Z}_p$ so that $|\alpha - k/p| \leq p^{-1}$, and it follows that, for any $j \neq k$, ki_j and ki_k are separated by at least $pn^{-2} - 3 \geq S$ integers modulo p . Write $W(k)$ for the k^{th} symbol of the word W . Then we may define $\chi: \mathbb{Z}_p \rightarrow [q]$ by $\chi(ki_j + t) = W_t(j)$ for $1 \leq t \leq S$ and $1 \leq j \leq n$, since the minimum gap between elements of $k\mathcal{I}$ is at least S . Define $\chi(s)$ arbitrarily for $s \neq ki_j + t$ for any j and t . Then the map $\chi': s \rightarrow \chi(ks)$ has the desired property. ■

We could use this result immediately to achieve a length $n^2 q^n(1 + o(1))$ Type I approximate \mathcal{I} -cycle, but it possible to do better using the random construction above. There is a trivial lower bound of q^n on the length of any approximate cycle, and we can show an upper bound that is only slightly worse:

Theorem 4.4. *For any $\mathcal{I} = \{i_j\}_{j=1}^n \subset \mathbb{Z}$, there exists a q -ary Type I approximate \mathcal{I} -cycle of length $p = (8 + o(1))q^n \log n$.*

Proof. The basic idea is to take a random sequence T_1 that contains almost all words, then use Lemma 4.3 to “tack on” the remaining ones. We use the notation of the proof of Theorem 4.1. If we let $F(n) = \lceil 4q^n \log n \rceil$, then, for sufficiently large n , the expected number of words which do not appear as $\chi(\mathcal{I} + t)$ is at most

$$q^n \exp(-2 \log n) = \frac{q^n}{n^2}.$$

Apply the lemma to find a sequence T_2 in which each word missed by the random sequence occurs. Then, we concatenate two copies of T_1 with two copies of T_2 (two copies are used to avoid disturbing sequences which “wrap around”), and the result is a Type I approximate \mathcal{I} -cycle of the stated length. ■

5. Decompositions into Directed Cycles

A number of recent papers (most notably [1], [2], and [3]) have addressed (and solved) the problem of decomposing a complete (possibly directed) graph into a set of cycles of prescribed length. Generally speaking, so long as the cycle lengths add up to the number of edges, there are very few impediments to the existence of such decompositions – although demonstrating this is far from simple. None of this work has dealt with graphs containing loops, however; in this section, we address this situation, in the case when all the cycles have the same length.

Let \tilde{K}_n denote the complete directed graph with loops on n vertices, i.e., the vertex set is $[n]$ and the edge set is $[n] \times [n]$, and let \overleftrightarrow{K}_n denote the complete directed graph without loops. We wish to know, for which n and d is it possible to decompose the edge set of \tilde{K}_n completely into cycles of length d ? Clearly d must divide n^2 . However, our main result states that the only other obstruction is that $d > 2$. (Indeed, it is easy to see in these cases that such a decomposition is not possible.)

First, we state a result from [3].

Theorem 5.1. *If $\sum_{i=1}^t m_i = n(n-1)$ and $m_i \geq 2$ for $i = 1, \dots, t$, then \overleftrightarrow{K}_n can be decomposed as the edge-disjoint union of cycles of lengths m_1, \dots, m_t , except in the case when $n = 6$ and all $m_i = 3$.*

This will imply the following:

Proposition 5.2. *If $d|n^2$, $d = 6$ or $d \geq 8$, then \tilde{K}_n may be decomposed into cycles of length d .*

Proof. We offer a procedure for “packing” length d cycles first into G , then the G_j ’s, then H . Split the vertex set of \tilde{K}_n into two pieces: $U = \{a, b\}$ and $V = \{1, \dots, n-2\}$. We may then decompose the edge set of \tilde{K} into the following pieces:

1. one \tilde{K}_2 on U ,
2. one $\overleftrightarrow{K}_{n-2}$ on V ,
3. and $n-2$ graphs each of which has vertex set $\{a, b, j\}$ for some $j \in V$, with a loop at j and edges (j, i) and (i, j) for $i = a, b$.

Such a decomposition is possible because $d \geq 6$ implies $n > 2$.

Call the first graph H , the second G , and the third G_j . Denote by $\{u, v\}$ the pair $\{(u, v), (v, u)\}$. Suppose $\binom{n-2}{2} \equiv r \pmod{d}$ with $1 \leq r \leq d$. Supposing $r > 1$, by Theorem 5.1, we may decompose G into $K = \lfloor (\binom{n-2}{2} - 1) / d \rfloor$ length d cycles and one r -trail, which we call T . We may assume, without loss of generality, that T meets vertex 3. Let $m = \lfloor (d-r)/5 \rfloor$ and $d' = d - r - 5m$. We construct a graph X as follows. Take X to be the union of T , G_j for $j = 1, \dots, m-1$, and one of the following, according to the value of d' . Note that $d' < 4$ implies $m \leq n-3$, since otherwise the number of remaining edges (i.e., ones unaccounted for thus far by X or any of the cycles in $\overleftrightarrow{K}_{n-2}$) would not be divisible by d . Similarly, $d' = 4$ implies either $m < n-3$ or we may add in all the remaining edges of the graph and be done.

- $d' = 0$: Add G_m to X .
- $d' = 1$: If $m = 0$, add the edge $(1, 1)$ to X . If $m > 0$, add the edges $\{a, m\}$, (m, m) , $\{a, m+1\}$, and $(m+1, m+1)$.
- $d' = 2$: Add $\{a, m+1\}$.
- $d' = 3$: Add $\{a, m+1\}$ and $(m+1, m+1)$.
- $d' = 4$: Add $\{a, m+1\}$ and $\{a, m+2\}$.

Note that in all cases, X is connected and has equal in- and out-degree at every vertex; therefore, X is Eulerian and may be written as a cycle of length d .

Now, if not all edges of \tilde{K}_n have been accounted for, yet, we wish to add another length d cycle. At the previous step, there are a few possibilities for the remaining set of edges not assigned to cycles. The set contains H , some of the G_j , as well as either:

1. Case I: $\{a, m\}$, $\{b, m\}$,
2. Case II: $\{b, m\}$, $\{b, m+1\}$,
3. Case III: (m, m) , $\{b, m\}$,
4. Case IV: $\{b, m\}$, or
5. Case V: (m, m) , $\{b, m\}$, $(m+1, m+1)$, $\{b, m+1\}$.

In each of the cases, let Y be the set of edges listed above. We now let $m = \lfloor (d - |Y|)/5 \rfloor$ and $d' = d - 5m - |Y|$. We may proceed exactly as above with the construction of X , unless $m = 0$. If $m = 0$, then either $d' \geq 2$ or $d' = 0$. In the latter case, $d = 6$ and we are in Case V, so we may simply take $X = Y$. In the former case, we append Y , as well as $d' > 0$ edges drawn from the first full “unused” G_k as follows. (Note that, if there is no unused G_k , we may append the remaining edges in the graph and be done.)

- $d' = 2$: Add $\{b, k\}$.
- $d' = 3$: Add $\{b, k\}$ and (k, k) .
- $d' = 4$: If $k < n - 2$, add $\{b, k\}$ and $\{b, k+1\}$. It is not possible for $k = n - 2$, since there would be too few edges left for d to divide them evenly.

Again, the resulting graph X is Eulerian, and can therefore be written as a length d cycle. It is clear that we may repeat the previous step until a full decomposition is achieved, though we may need to switch the roles of a and b .

It remains to deal with the case of $r = 1$ at the beginning of the proof. Instead of decomposing into all length d trails except for one r -trail, we instead create one length $d - 1$ trail and one length 2 trail. Then attach (a, a) to the $d - 1$ trail (changing labels if necessary to maintain connectivity), and proceed as above with $r = 2$, using the remaining edges of H when necessary. All of the details of the preceding argument work with this slight modification. ■

Proposition 5.3. *If n is even, then \tilde{K}_n may be decomposed into cycles of length 4.*

Proof. There are two cases: $n \equiv 2 \pmod{4}$ and $n \equiv 0 \pmod{4}$. Suppose the former. Then the following types of edge-sets partition $E(\tilde{K}_n)$ into cycles of length 4:

- (1) $\{(j, j), (j + n/2, j + n/2), (j, j + n/2), (j + n/2, j)\}$ for each j with $1 \leq j \leq n/2$,
- (2) $\{(j, j + 2k - 1), (j + 2k - 1, j), (j, j + 2k), (j + 2k, j)\}$, for each j and each k so that $1 \leq k \leq (n - 2)/4$.

If $4|n$, we use the following types of sets instead:

- (1) $\{(j, j), (j + n/2, j + n/2), (j, j + n/2), (j + n/2, j)\}$ for each j with $1 \leq j \leq n/2$,
- (2) $\{(j, j + 2k), (j + 2k, j), (j, j + 2k + 1), (j + 2k + 1, j)\}$, for each j and each k so that $1 \leq k \leq n/4 - 1$,
- (3) $\{(2j, 2j - 1), (2j - 1, 2j), (2j, 2j + 1), (2j + 1, 2j)\}$, for each j with $1 \leq j \leq n/2$.

■

Proposition 5.4. *If $d|n$, $d = 3, 5$, or 7 , then \tilde{K}_n may be decomposed into cycles of length d .*

Proof. We imitate the case of $d \geq 8$ here, only the situation is simpler. Suppose $d = 3$. We may decompose \tilde{K}_n into three pieces: the loop (a, a) , a \overrightarrow{K}_{n-1} , and $n - 1$ length 3 cycles G_1, \dots, G_{n-1} . Applying Theorem 5.1, we may break the second of these into cycles of length 3, except for one of length 2, which we call T . (Since $3|n^2$, $(n - 1) \cdot (n - 2) \equiv 2 \pmod{3}$.) We may assume that T meets a ; therefore, appending (a, a) to T and taking each G_j as its own cycle provides a decomposition.

Now suppose $d = 5$. We may decompose \tilde{K}_n into three pieces again: a \tilde{K}_2 on $\{a, b\}$, a \overrightarrow{K}_{n-2} , and $n - 2$ length 5 cycles G_1, \dots, G_{n-2} . Applying Theorem 5.1, we may break the \overrightarrow{K}_{n-2} into cycles of length 5, except for one of length 4 and one of length 2, which we call T_1 and T_2 , respectively. (This time, $(n - 2)(n - 3) \equiv 1 \pmod{5}$.) We may assume that T_1 meets a and T_2 meets b . Append (a, a) to T_1 , $\tilde{K}_2 \setminus (a, a)$ to T_2 , and take each G_j as its own element of the decomposition.

Finally, let $d = 7$. We have the following decomposition of \tilde{K}_n : a \tilde{K}_3 on $\{a, b, c\}$, a \overrightarrow{K}_{n-3} , and $n - 3$ length 7 cycles G_1, \dots, G_{n-3} . Applying Theorem 5.1, we may break the \overrightarrow{K}_{n-3} into cycles of length 7, except for one of length 5, which we call T' . (Now, $(n - 3)(n - 4) \equiv 5 \pmod{7}$.) We may assume that T' meets a . Append $\{a, b\}$ to T' , include the 7-cycle $\tilde{K}_3 \setminus \{a, b\}$, and take each G_j as its own element of the decomposition. ■

All of these results together imply the following.

Theorem 5.5. *If $d|n^2$ and $d \geq 3$, then \tilde{K}_n may be decomposed into cycles of length d .*

6. Conclusion

We wish to know, first and foremost, what distinguishes valid sets from invalid ones. The authors' attempts to find a simple way to separate these cases was met with frustration. A simpler problem is the case of index sets which are arithmetic progressions: the question of decomposing de Bruijn graphs into cycles is a natural one. We would also like to see corresponding lower bounds or improvements on the bounds of Section 4, particularly in the case of Type I approximate cycles.

Finally, the problem solved by Theorem 5.5 has a natural generalization along the lines of other, similar work. Suppose that $\{m_j\}$ is such that $\sum_j m_j = n^2$. When is it possible to decompose \tilde{K}_n into cycles of lengths $\{m_j\}$? Clearly, there can be at most n 1's among the m_j , and there must be at least n indices j for which $m_j \neq 2$. Are there additional restrictions?

Acknowledgment. Thanks to Paul Balister for the helpful suggestion concerning Theorem 5.5.

References

1. B. Alspach, H. Gavlas, M. Sajna, and H. Verrall, Cycle decompositions IV, Complete directed graphs and fixed length directed cycles, *J. Combin. Theory, Ser. A* **103** (1) (2003) 165–208.
2. P. Balister, Packing circuits into K_N , *Combin. Probab. Comput.* **10** (6) (2001) 463–499.
3. P. Balister, Packing digraphs with directed closed trails, *Combin. Probab. Comput.* **12** (1) (2003) 1–15.
4. T.B. Beard Jr., Matrix fields, regular and irregular: a complete fundamental characterization, *Linear Algebra Appl.* **81** (1986) 137–152.
5. F. Chung and J.N. Cooper, De Bruijn cycles for covering codes, preprint.
6. H. Fredricksen, A survey of full length nonlinear shift register cycle algorithms, *SIAM Rev.* **24** (2) (1982) 195–221.
7. S. Janson, New versions of Suens correlation inequality, *Proceedings of the Eighth International Conference, Random Structures Algorithms* **13** (3-4) (1998) 467–483.
8. S.V. Konyagin, I.Z. Rusza, and W. Schlag, On uniformly distributed dilates of finite integer sequences, *J. Number Theory* **82** (2000) 165–187.
9. R. Lidl and H. Niederreiter, *Finite Fields, Encyclopedia of Mathematics and Its Applications* **20**, Cambridge University Press, Cambridge, 1997.