CONSTRUCTION AND OPTIMALITY OF UNORIENTED DE BRUIJN SEQUENCES

CHRISTIE BURRIS
COLORADO STATE UNIVERSITY

ORIGINAL DE BRUIJN SEQUENCES

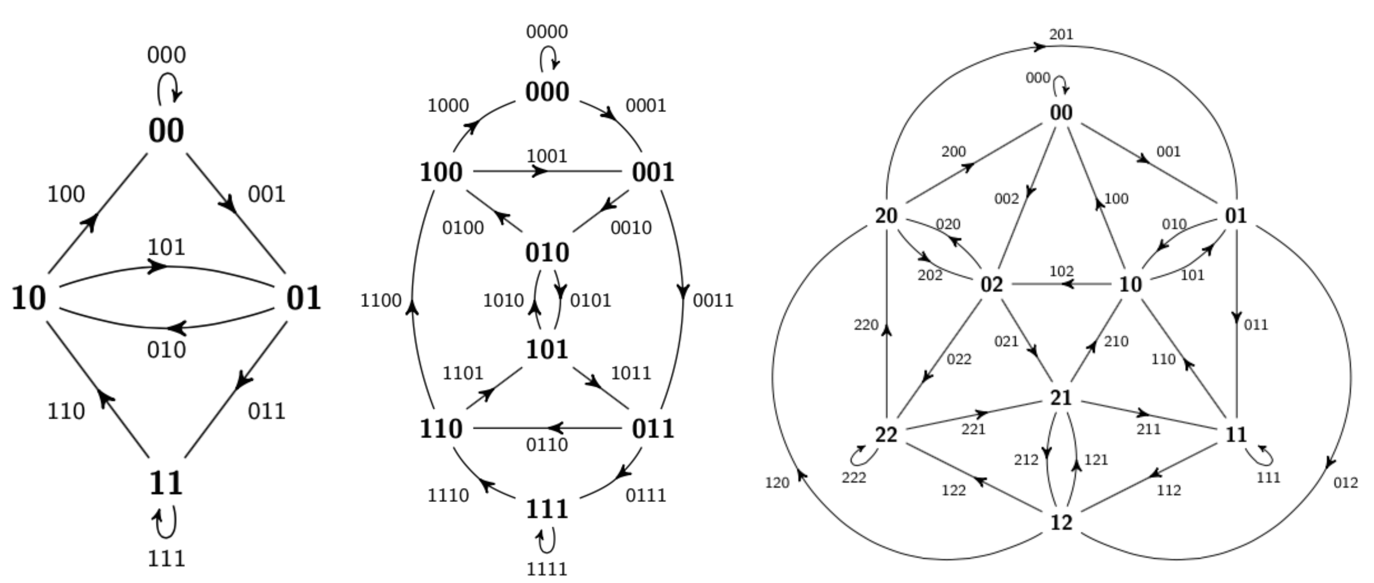
For positive integers k and n, does there exist a word which has as subwords every word of length n on k symbols exactly once?

Yes; a de Bruijn sequence $B(k,n)$! Its length-n subwords are exactly the set of k^n words of length n on k symbols. The length of such a sequence is $k^n + n - 1$. Consider the example where $k = 2$, $n = 3$.

$$B(2,3) = \mathbf{0100011101}$$

$B(2,3)$ contains every binary word of length 3—010, 100, 000, 001, 011, 111, 110, 101—exactly once.

Constructing de Bruijn sequences from de Bruijn graphs Bg(k,n)



(a) $Bg(2,3)$ (b) $Bg(2,4)$ (c) $Bg(3,3)$

Figure 1: $Bg(k,n)$ with vertices corresponding to words of length $(n-1)$ on k symbols. Each edge connects a word to each possible consecutive word in a sequence. Beginning at any vertex, $B(k,n)$ can be formed by tracing an Eulerian circuit—a path that sees every edge exactly once, and begins and ends at the same vertex.

In the case of Fig. 1a, the sequence 0100011101 is formed by the Eulerian circuit

$$01 \to 10 \to 00 \to 00 \to 01 \to 11 \to 11 \to 10 \to 01$$

which corresponds to the subwords in the order seen above

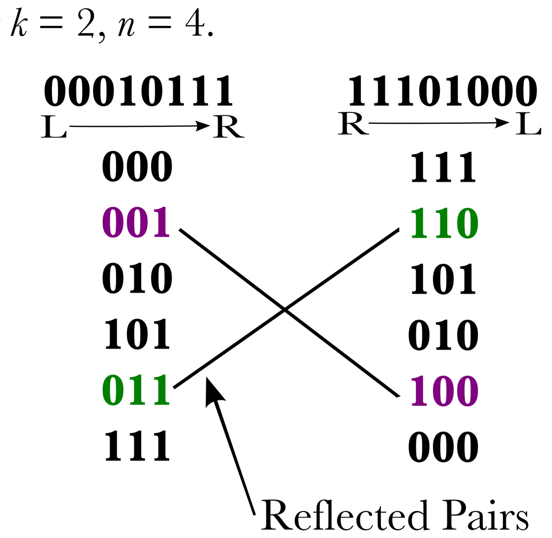
$$010 \to 100 \to 000 \to 001 \to 011 \to 111 \to 110 \to 101.$$

UNORIENTED DE BRUIJN SEQUENCES

Consider the sequence 00010111. This sequence has each of the binary palindromes of length 3 as subwords, 000, 010, 101, and 111, but only one member of each of the pairs {001, 100} and {011, 110}. Each non-palindrome word of length 3 appears either forwards or backwards exactly once. We call such a sequence an *unoriented de Bruijn sequence of optimal length*. Two words $\mathbf{v}$ and $\mathbf{v'}$ are *reflections or reflected pairs* if $\mathbf{v} = v_1 v_2 \ldots v_{n-1} v_n$ and $\mathbf{v'} = v_n v_{n-1} \ldots v_2 v_1$. Table 1 provides the reflected pairs for $k = 2$, $n = 4$.

Table 1: Reflected Pairs: k = 2, n = 4

w	**w'**	**w = w'**
0001	1000	0000
0010	0100	0110
0011	1100	1001
0101	1010	1111
0111	1110	
1011	1101	



Definition An *unoriented de Bruijn sequence uB(k,n)* is a sequence of characters drawn from an alphabet Σ_k of k symbols that (i) contains as subwords a member from each of the length-n reflections on k symbols, and (ii) is of the shortest length amongst all such sequences satisfying property (i).

When read forward and backward, the $uB(k,n)$ sequence sees every word of length n on k symbols. We say that an *unoriented de Bruijn sequence of optimal length* satisfies the additional property that (iii) it contains as subwords exactly one member from each of the length-n reflections on k symbols.

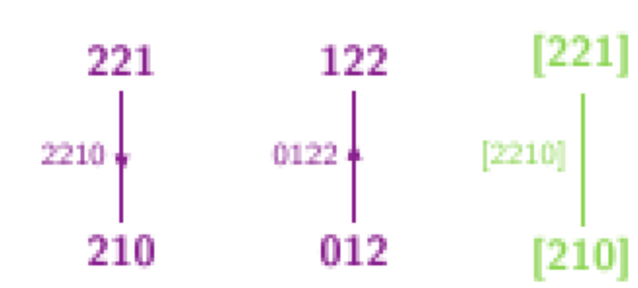
CONSTRUCTION & COMPLEXITY

Forming unoriented de Bruijn sequences

As an original de Bruijn sequence is formed following an Eulerian circuit in a de Bruijn graph, an unoriented de Bruijn sequence of optimal length would be formed by following a path in a de Bruijn graph that traverses exactly one of the two edges corresponding to each reflected pair.

Complexity of unoriented de Bruijn graphs

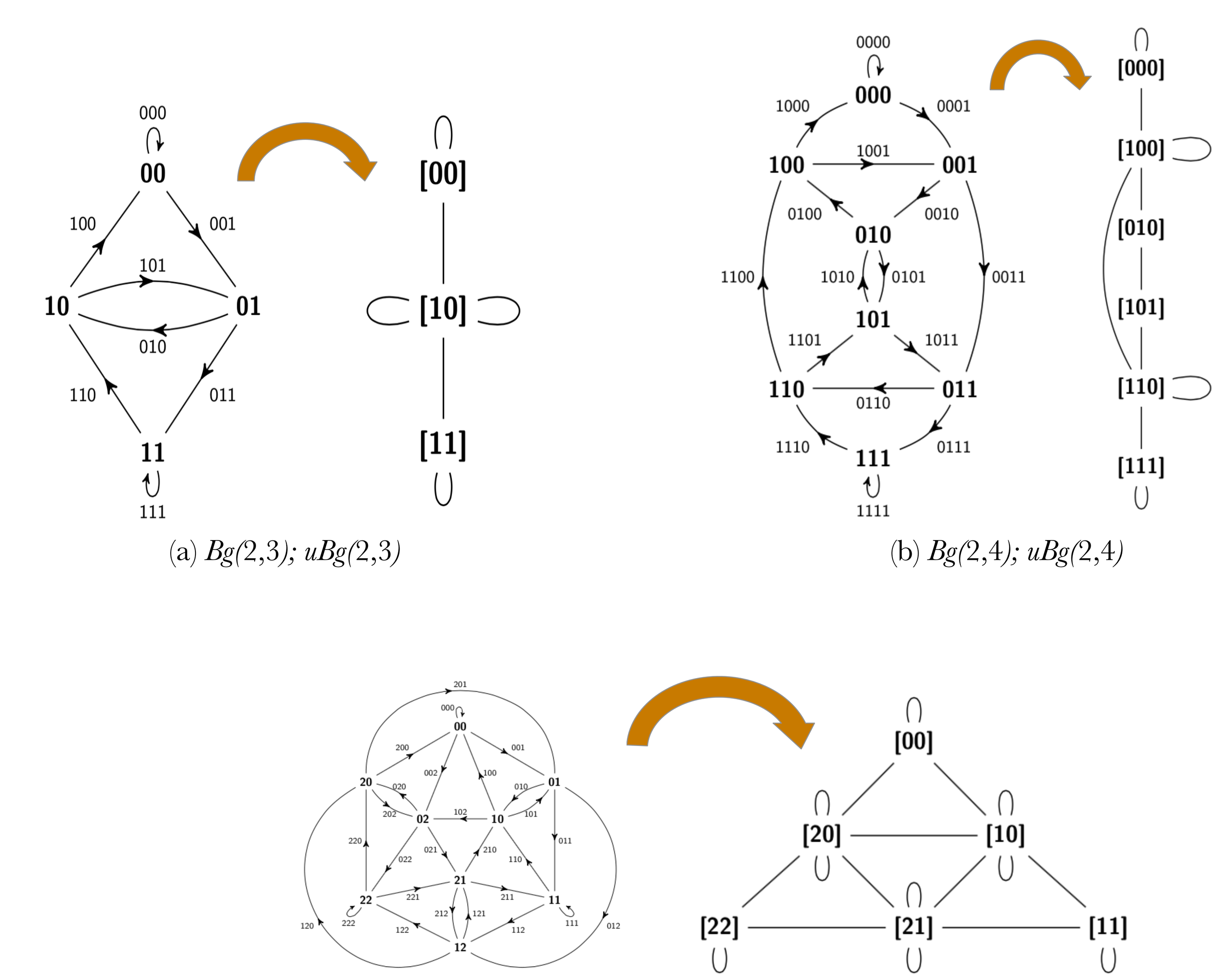
- Edges in the (k,n)-unoriented de Bruijn graph are identified by the pair of reflections that can be formed by the adjacent vertices. The case below holds in general.



- As the graph is traversed, the subword that is seen in the sequence is determined by the direction taken on each edge. As such, the ability to traverse an edge is conditional upon which of the two length-$(n-1)$ words was previously traversed.

$$[11] \to [21] \to [21] \to [21] \to [22] \qquad [11] \to [21] \to [21] \not\to [22]$$
$$112 \to 121 \to 212 \to 122 \qquad 112 \to 121 \not\to 122$$

- In order to form a valid unoriented de Bruijn sequence, one must traverse an *alternating Eulerian path* – an Eulerian path that respects the above conditions.



(a) $Bg(2,3)$; $uBg(2,3)$ (b) $Bg(2,4)$; $uBg(2,4)$ (c) $Bg(3,3)$; $uBg(3,3)$

Figure 2: Original de Bruijn graphs (left) and analogous unoriented de Bruijn graphs (right) which contain $(k^n+k^{n/2})/2$ vertices labelled by the equivalence classes $[\mathbf{v}]$ of k-ary words of length $(n-1)$. There is an (undirected) edge $[\mathbf{v}] \to [\mathbf{w}]$ if $\mathbf{v} \to \mathbf{w}$ or $\mathbf{v} \to \mathbf{w'}$ is a (directed) edge in the de Bruijn graph $Bg(k,n)$.

Length of unoriented de Bruijn sequences

The length of any optimal $uB(k,n)$ sequence $L(k,n)$, is the length required to see all palindromes of length n on k symbols and half of all non-palindromes.

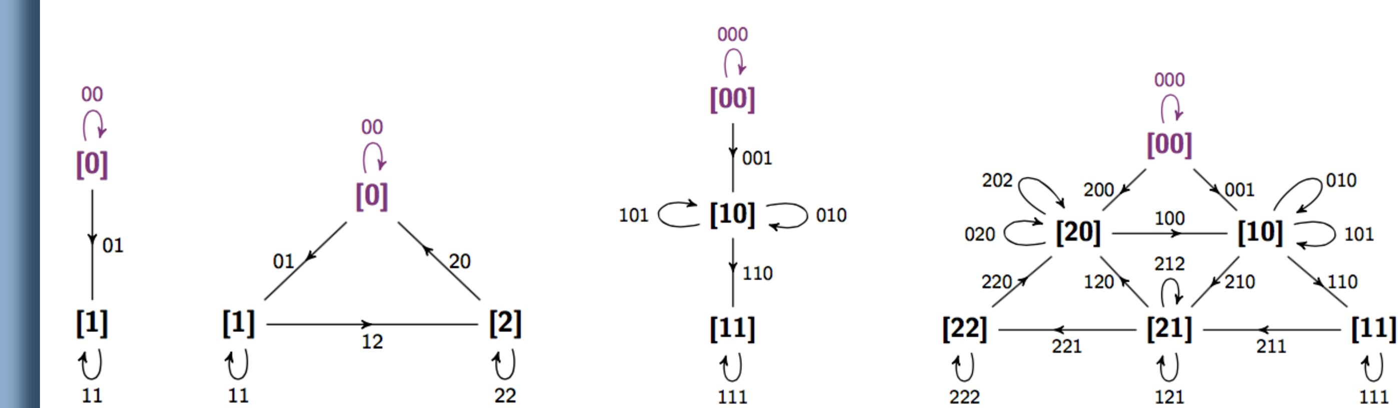
$$L(k,n) = (k^n + k^{\lceil n/2 \rceil} + 2n - 2)/2.$$

EXISTENCE

Theorem 1. An unoriented de Bruijn sequence of optimal length $uB(k,n)$ exists iff k is 2 or odd and $n \leq 3$.

Table 2: Optimal Unoriented de Bruijn Sequences

	$n = 2$	$n = 3$
$k = 2$	0011	00010111
$k = 3$	0011220	0001011121202012 2200



(a) $uBg(2,2)$ (b) $uBg(3,2)$ (c) $uBg(2,3)$ (d) $uB(3,3)$

Figure 3: Optimal unoriented de Bruijn sequences represented on unoriented de Bruijn graphs. Note that unoriented de Bruijn graphs are undirected. The paths taken on each graph (a) - (d) follow the directed edges, beginning at the constant 0 vertex (colored red).

The proof is constructed by counting $ov(k,n)$, the number of odd-degree vertices in unoriented de Bruijn graphs. Only vertices for which exactly one of either the first or last $(n-2)$ characters form a palindrome have odd degree. The count of such vertices is found by subtracting from (total #-vertices) the (#-vertices with the above property). For k even and greater than 2 and n greater than 3, there are more odd vertices than are allowed.

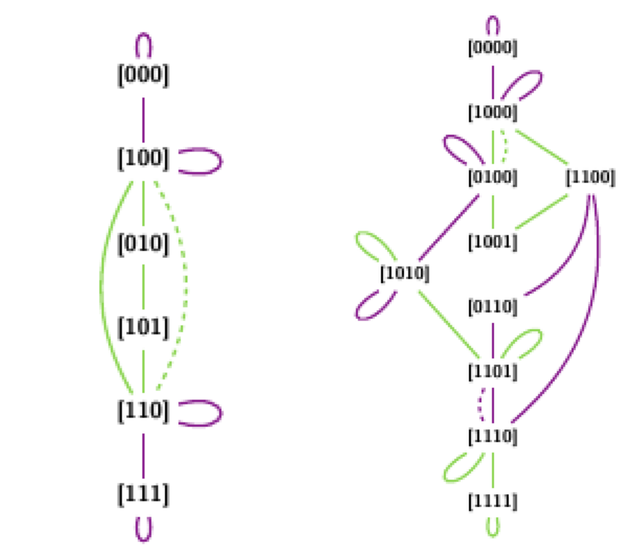
FUTURE WORK

uB(k,n) of suboptimal length

What is the minimum length of existing unoriented de Bruijn sequences where either k or n is greater than 3? These sequences would be constructed in such a way that the minimum number of edges are repeated in $uBg(k,n)$ in order for the resulting graph to have an alternating Eulerian path, as in Fig. 5.

(a) **00001011001111**

(b) **0000010001110110010101 1111**



(a) $uBg(2,4)$ (b) $uBg(2,5)$

Figure 5 (top): Unoriented de Bruijn graphs with additional edges (dotted) that give rise to unoriented de Bruijn sequences of suboptimal length.

How optimal is suboptimal?

The length of an unoriented de Bruijn sequence is bounded above by $L(k,n) + (n-1)(ov(k,n)/2 - 1)$.

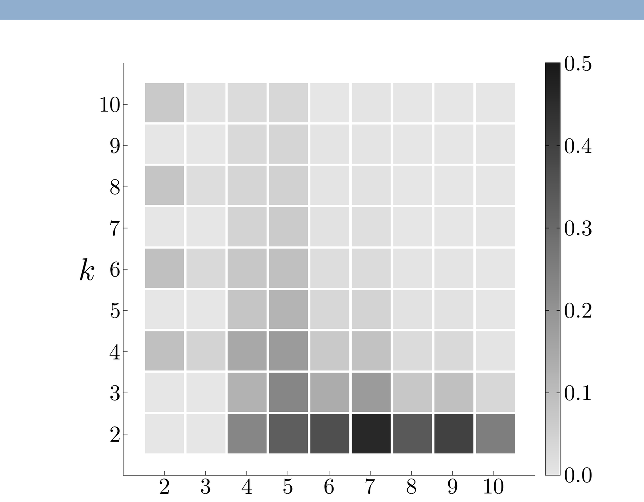


Figure 6 (left): Plot of $r(k,n)$, the ratio of the upper bound on the length of $uB(k,n)$ to the length of $B(k,n)$ over the range $2 \leq n, k \leq 10$. Note the rapid convergence to 1/2 as n or k increases.

This work is supported by the NSF grant DMS- 1022635 to Patrick Shipman.