ON THE EXISTENCE OF BALANCED GENERALIZED DE BRUJN SEQUENCES

Matthew Baker  
School of Mathematics  
Georgia Institute of Technology  
Atlanta, GA 30332  
mbaker@math.gatech.edu

Bhumika Mittal  
Ashoka University  
Sonipat, Haryana 131029  
bhumika.mittal_ug24@ashoka.edu.in

Haran Mouli  
PSBB KKN  
Chennai, Tamil Nadu 600078  
mouliharan@gmail.com

Eric Tang  
Harvard University  
Cambridge, MA 02138  
etang@college.harvard.edu

January 31, 2022

ABSTRACT

A balanced generalized de Bruijn sequence with parameters \((n, l, k)\) is a cyclic sequence of \(n\) bits such that (a) the number of 0’s equals the number of 1’s, and (b) each substring of length \(l\) occurs at most \(k\) times. We determine necessary and sufficient conditions on \(n, l,\) and \(k\) for the existence of such a sequence.

1 Statement of the main theorem

De Bruijn sequences, named after Nicolaas Govert de Bruijn (who first wrote about them in 1946) but first explored systematically by Camille Flye Sainte-Marie in 1894, are well-studied in mathematical literature. A (binary) de Bruijn sequence of order \(m\) is a cyclic sequence where every possible \(m\)-bit substring occurs exactly once. It is well known ([dB46], see also [Hal86]) that there are \(2^{2^{m-1}-m}\) distinct de Bruijn sequences of order \(m\). In this paper, we generalize de Bruijn sequences in the following manner:

Definition 1. A generalized de Bruijn sequence with parameters \((n, l, k)\) is a cyclic sequence of \(n\) bits such that each substring of length \(l\) occurs at most \(k\) times. Such a sequence is called balanced if the number of 0s and 1s are equal.

In our terminology, a (classical) de Bruijn sequence of order \(m\) is a generalized de Bruijn sequence with parameters \((2^m, m, 1)\). Such a sequence is automatically balanced.

The main result of this paper characterizes those parameters \((n, l, k)\) for which a balanced generalized de Bruijn sequence with parameters \((n, l, k)\) exists:

Theorem 2. Given positive integers \(n, l,\) and \(k,\) a balanced generalized de Bruijn sequence with parameters \((n, l, k)\) exists if and only if \(n\) is even and \(k \geq \frac{n}{2}.

*The second, third, and fourth authors completed this research under the direction of the first author as part of the summer 2021 PROMYS program. We thank the PROMYS program for their support and for creating such a stimulating work environment. We also thank Persi Diaconis for helpful feedback on an earlier draft. The first author was supported by a Simons Foundation Collaboration Grant.
In Section 3 we discuss our original motivation for introducing balanced generalized de Bruijn sequences, which came from the world of mathematical card tricks.

2 Proof of the main theorem

Our proof of Theorem 2 used ideas from graph theory. Specifically, we will be concerned with certain subgraphs of the following family of directed graphs:

Definition 3. For \( l \geq 2 \), the de Bruijn graph of rank \( l \), denoted by \( G_l \), is a directed graph on \( 2^{l-1} \) vertices, labeled by the elements of \( \{0, 1\}^{l-1} \), and \( 2^l \) edges, labeled by the elements of \( \{0, 1\}^l \). For any \( l \) bits \( b_0, b_1, \cdots, b_{l-1} \), the edge \( e = b_0b_1 \cdots b_{l-1} \) connects \( v_1 = b_0b_1 \cdots b_{l-2} \) to \( v_2 = b_1b_2 \cdots b_{l-1} \).

We begin with some simple lemmas.

Lemma 4. Every vertex in the de Bruijn graph \( G_l \) has both in-degree and out-degree 2.

Proof. Consider the vertex \( v = b_0b_1 \cdots b_{l-2} \in V(G_l) \). By inspection, \( v \) has in-edges \( bb_0b_1 \cdots b_{l-2} \) and out-edges \( b_0b_1 \cdots b_{l-2}b \) where \( b \in \{0, 1\} \). Thus, \( v \) has both in-degree and out-degree 2.

Lemma 5. The de Bruijn graph \( G_l \) is Eulerian.

Proof. We know from Lemma 4 that every vertex in \( G_l \) has equal in-degree and out-degree. Thus, by a well-known criterion for the existence of Eulerian circuits in digraphs (see for example [Bru10, Theorem 13.1.2]) it suffices to show that \( G_l \) is connected. This follows from the fact that given vertices \( v = b_0b_1 \cdots b_{l-2} \) and \( v' = b'_0b'_1 \cdots b'_{l-2} \), the following is a walk from \( v \) to \( v' \):

\[
\begin{align*}
&b_0b_1 \cdots b_{l-2} \to b_1 \cdots b_{l-2}b'_0 \to \cdots \to b'_0b'_1 \cdots b'_{l-2}.
\end{align*}
\]

It will be convenient to color each edge of \( G_l \) either red or blue according to the following rule:

Definition 6. We define an edge in \( G_l \) to be red if its last bit is 0 and blue if its last bit is 1. A subgraph of \( G_l \) is balanced if it contains an equal number of red and blue edges.

Lemma 7. Each vertex \( v \in V(G_l) \) has one red and one blue out-edge, and the two in-edges of \( v \) have the same color.

Proof. Let \( v = b_0b_1 \cdots b_{l-2} \) be a vertex in \( G_l \). Then by definition, the out-edges \( b_0b_1 \cdots b_{l-2}0 \) and \( b_0b_1 \cdots b_{l-2}1 \) are red and blue, respectively. The in-edges \( 0b_0b_1 \cdots b_{l-2} \) and \( 1b_0b_1 \cdots b_{l-2} \) must end with the same bit, and are thus of the same color.

Recall that a circuit in a graph is a path which begins and ends at the same vertex, and a cycle is a circuit in which no vertex (other than the initial and final one) is repeated.

By symmetry, the deBruijn graph \( G_l \) itself is balanced for all \( l \geq 2 \). We are interested in certain balanced subgraphs of \( G_l \) because of the following simple observation:

Lemma 8. There is a one-to-one correspondence between balanced circuits of length \( n \) in \( G_l \) and balanced generalized de Bruijn sequences with parameters \((n, l, 1)\).

Proof. Let the edges of the circuit of length \( n \) be denoted as:

\[
\begin{align*}
b_0b_1 \cdots b_{l-1} \to b_1b_2 \cdots b_l \to \cdots \to b_{n-1}b_0 \cdots b_{l-2}.
\end{align*}
\]

The \( l \)-bit substrings of the \( n \)-bit sequence \( b_0b_1 \cdots b_{n-1} \) precisely correspond to the edges of the circuit of length \( n \). Thus, as the edges on the circuit are distinct, so are the \( l \)-bit substrings. Moreover, since the number of red and blue edges in the circuit are equal, the number of 0s and 1s in the string are equal as well. Hence, the \( n \)-bit string is a balanced generalized de Bruijn sequence with parameters \((n, l, 1)\).

Lemma 9. If the graph \( G_l \) has a balanced circuit of length \( n \), then the graph \( G_{l+1} \) has a balanced cycle of length \( n \).
Proof. The edges of $G_i$ and vertices of $G_{i+1}$ have the same labelling. This gives us a natural bijection between the edges of $G_i$ and the vertices of $G_{i+1}$. Let $e_1, e_2 \in G_i$ correspond to $v_1, v_2 \in G_{i+1}$ under this bijection. It follows from the definition of a de Bruijn graph that the head of $e_1$ is the tip of $e_2$ in $G_i$ if and only if there is an edge connecting $v_1$ to $v_2$ in $G_{i+1}$. Now, consider the edges of the circuit of length $n$ in $G_i$:

$$b_0b_1\cdots b_{l-1} \rightarrow b_1b_2\cdots b_l \rightarrow \cdots \rightarrow b_{n-1}b_0\cdots b_{l-2}.$$ 

We consider these edges as the vertices of a walk in $G_{i+1}$. As all the edges of the circuit are distinct, the walk in $G_{i+1}$ contains distinct vertices. Thus, the bijection maps a circuit of length $n$ in $G_i$ to an $n$-cycle in $G_{i+1}$. Moreover, since the circuit is balanced, there are equal number of 0s and 1s among $b_0, b_1, \cdots, b_{n-1}$. These will be the last bits in the edges of the cycle in $G_{i+1}$ as well, and hence, the cycle is balanced.

Lemma 10. Let $l \geq 2$, and let $n$ be an even positive integer at most $2^l$. Then the graph $G_i$ contains a balanced circuit of length $n$.

Proof. We proceed by induction on $l$. First, for $l = 2$, we can either have $n = 2$ or $n = 4$. We may take the circuits $0 \rightarrow 1$ and $0 \rightarrow 1 \rightarrow 0$, respectively.

Now, assume that the claim holds true for $l$. We prove it holds true for $l+1$. For even $n \leq 2^l$, we may use the induction hypothesis to find a balanced circuit of length $n$ in $G_i$. From Lemma 9 there exists a balanced cycle (also a circuit) of length $n$ in $G_{i+1}$.

We now show the claim for $2^l < n < 2^{l+1}$ where $n$ is even. Since $2^l \leq 2^{l+1} - n < 2^l$, there exists a balanced $n$-cycle in $G_{i+1}$, say $H$. Then the graph $G_{i+1} - H$ contains $n$ edges. Assume that $G_{i+1} - H = H_1 \cup \cdots \cup H_t$, where each $H_i$ for $1 \leq i \leq t$ is a component of $G_{i+1} - H$. Since both $G_{i+1}$ and $H$ are balanced and Eulerian, $G_{i+1} - H$ must be balanced and each $H_i$ must be Eulerian.

If $G_{i+1} - H$ is connected, we are done, since it is a circuit in $G_{i+1}$ of length $n$. Otherwise, since $G_{i+1}$ is connected, there exists an edge $e$ in $H$ which connects vertices in two different components of $G_{i+1} - H$, say $H_1$ and $H_2$. Let the edge $e$ go from $v_1 \in V(H_1)$ to $v_2 \in V(H_2)$. Without loss of generality, assume $e$ is red. Now, let $e_1$ be an edge from $v_1$ to $u_1$ in $H_1$, and let $e_2$ be an edge from $u_2$ to $v_2$ in $H_2$. Then the last $l-1$ bits of $v_1$ and $u_1$. Hence, there must be an edge $e'$ from $u_2$ to $v_1$ in $H$. We deduce from Lemma 9 that since $e$ is red, $e_1$ must be blue, $e_2$ must be red, and $e'$ must be blue.

Consider the new subgraph obtained by replacing $e_1$ and $e_2$ with $e$ and $e'$. One checks easily that the degrees of all vertices, as well as the number of red and blue edges, are preserved. Moreover, the number of connected components has been reduced. Thus, we end up with a balanced subgraph with $n$ edges and fewer components, all of which are Eulerian. By repeating this process, we will eventually end up with a balanced subgraph of $n$ edges which is a circuit. Finally, for $n = 2^{l+1}$, all of the edges of the graph form a circuit since $G_{i+1}$ is Eulerian. The result follows.

With these preliminaries out of the way, we now turn to the proof of our main result:

Proof of Theorem 2. ($\Rightarrow$) Since a balanced generalized de Bruijn sequence has an equal number of 0s and 1s, $n$ must be even. Moreover, by the Pigeonhole Principle, we must have $k \geq \frac{n}{2^l}$, since there are only $2^l$ possible $l$-bit substrings and $n$ total substrings.

($\Leftarrow$) Now, we show that these constraints are sufficient for a balanced generalized de Bruijn sequence to exist. We may assume that $k = \lceil \frac{n}{2^l} \rceil$ since a balanced generalized de Bruijn sequence with parameters $(n, l, k)$ is also one with parameters $(n, l, k')$ where $k' > k$.

For $l = 1$, the claim is trivial since the substrings are simply bits, and we may use any balanced string of length $n$. Now, let $l \geq 2$. We will now prove the claim by induction on $k$. The base case $k = 1$ follows from Lemmas 8 and 10.

It suffices to show that the existence of a balanced generalized de Bruijn sequence with parameters $(n, l, k)$ implies the existence of one with parameters $(n + 2^l, l, k + 1)$. Let $b_0b_1\cdots b_{n-1}$ be a balanced generalized de Bruijn sequence with parameters $(n, l, k)$, and let $b'_0b'_1\cdots b'_{2^n-1}$ be one with parameters $(2^l, l, 1)$. The latter contains every $l$-bit string as a substring exactly once. We may therefore assume that $b'_0b'_1\cdots b'_{2^l-1} = b_0b_1\cdots b_{l-1}$. Thus

$$b_0b_1\cdots b_{n-1}b'_0b'_1\cdots b'_{2^l-1}$$

is a balanced sequence in which each substring of length $l$ appears at most $k$ times in the first part of the sequence and exactly once in the second part. It is therefore a balanced generalized de Bruijn sequence with parameters $(n + 2^l, l, k + 1)$. The theorem follows by induction.
Remark 11. Theorem 2 can be generalized as follows. Say that a (finite) binary sequence is almost-balanced if the number of 0’s and the number of 1’s differ by at most one. Since the complement of an almost-balanced sequence is also almost-balanced, the same argument used to prove Theorem 2 yields the following:

**Theorem 12.** Given positive integers \( n, l, \) and \( k \), an almost-balanced generalized de Bruijn sequence with parameters \( (n, l, k) \) exists if and only if \( k \geq \frac{n}{2} \).

### 3 Motivation

One of our motivations for introducing balanced generalized de Bruijn sequences comes from the following card trick, a 32-card version of which is described on p.18 of [DG12]. The 52-card variant described here is alluded to on p.29 of loc. cit.

**Effect and presentation.**

A regular deck of 52 cards, secured with rubber bands in its case, is tossed out into the audience. The deck is tossed around a few times until finally one audience member is instructed to remove the rubber bands, open the case, and give the deck a cut at some random position. This audience member takes the top card and then passes it to another spectator who removes the current top card. This is repeated until five members of the audience have taken cards. The performer instructs the five participants who have taken cards to look at their card, make a mental picture of it, and try to transmit an image of their card to the performer. The performer says, “You’re all doing great, but some of your signals are crossing and I can’t quite make everything out. Would anyone who has a red card please stand up and try extra hard to project the mental image of their card?” Turning to the first participant that took a card, the performer says, “I’m having trouble making out the suit of your card... it’s not a heart, is it?” The participant shakes his head no. “No, I didn’t think so... no, it’s a diamond. It’s the 9 of Diamonds!” The participant nods in agreement. Without asking any more questions, the performer goes on to identify the exact card that each of the other four participants is thinking of.

**Method and explanation.**

The trick is based on a balanced generalized de Bruijn sequence with parameters \( (52, 5, 2) \), for example the following one:

\[
00000 : A\spadesuit, A\heartsuit \quad 01000 : 10\spadesuit, K\spadesuit 10000 : 4\spadesuit, 4\heartsuit 11000 : 7\spadesuit \\
00001 : 7\spadesuit, 7\heartsuit \quad 01001 : K\spadesuit, 5\spadesuit 10001 : 6\spadesuit 11001 : J\spadesuit \\
00100 : 6\spadesuit, 6\heartsuit \quad 01100 : 2\spadesuit \quad 10100 : 3\spadesuit, A\spadesuit 11100 : 8\spadesuit, J\spadesuit \\
00101 : 3\spadesuit \quad 01101 : 9\spadesuit, 9\spadesuit 10111 : 5\spadesuit 11111 : A\spadesuit, 2\spadesuit \\
01001 : 5\spadesuit, 10\spadesuit \quad 01110 : J\spadesuit \quad 10101 : 7\spadesuit, 3\spadesuit 11101 : 10\spadesuit \\
01101 : 8\spadesuit, 8\heartsuit \quad 01101 : Q\spadesuit, 4\spadesuit 10101 : 10\spadesuit 11110 : K\spadesuit, 8\spadesuit \\
01110 : 3\spadesuit \quad 01111 : 2\spadesuit 10110 : Q\spadesuit, Q\spadesuit 11110 : 9\spadesuit, K\spadesuit \\
01111 : Q\spadesuit \quad 01111 : 4\spadesuit, J\spadesuit 10111 : 6\spadesuit, 2\spadesuit 11111 : 5\spadesuit, 9\spadesuit
\]

If there are two possibilities listed, for example 9H or 9D, the magician feints hesitation, saying (in this example) to the first spectator, “I’m having trouble seeing the suit... it’s not a heart, is it?” If the spectator says yes, the magician responds, “Yes, I thought so! It’s the 9 of Hearts!” If the spectator says no, the magician responds, “No, I didn’t think so... no, it’s a diamond. It’s the 9 of Diamonds!” The participant nods in agreement. Without asking any more questions, the performer goes on to identify the exact card that each of the other four participants is thinking of.

To disguise the glimpsing of the necessary information, once the spectators have signaled the color of their cards, grab a clipboard and a pen and say that you’re going to try to write down the exact cards held by each of the five participants. Unbeknownst to the participants, on the clipboard you have a crib sheet containing the above information.
4 Some open problems

1. In lieu of binary bits, consider an alphabet $A$ of size $m$ where $m > 2$. Does a suitable analogue of Theorem 2 still hold? That is, given positive integers $n, l, k,$ and $m$, does an $A$-balanced generalized deBruijn sequence of length $n$ exist if and only if $m$ divides $n$ and $k \geq \frac{n}{m}$? As considering the complement of such a sequence no longer suffices, the induction step used to prove Theorem 2 does not apply in this more general context.

2. Suppose $n$ is even and $k \geq \frac{n}{2}$. How many balanced generalized de Bruijn sequences with parameters $(n, l, k)$ are there? We know there is at least 1, and when $(n, l, k) = (2^m, m, 1)$ there are precisely $2^{2m-1}-m$ such sequences. Is there an easily described general formula – at least in some special cases?

3. Are there algebraic ‘shift register’ type constructions that would permit the performer to do the (52, 5, 2) trick described in Section 3 without mnemonics or a list? The trick described on p.18 of [DG12] (with parameters (32, 5, 1)) has this property. For analogous constructions related to a 6-card version of our (52, 5, 2) trick, see [LP12] or [Son].

References

[Bru10] Richard A. Brualdi. *Introductory Combinatorics*. Pearson Prentice Hall, Upper Saddle River, NJ, fifth edition, 2010.

[dB46] N. G. de Bruijn. A Combinatorial Problem. *Nederl. Akad. Wetensch., Proc.*, 49:758–764 = Indagationes Math. 8, 461–467 (1946), 1946.

[DG12] Persi Diaconis and Ron Graham. *Magical Mathematics*. Princeton University Press, Princeton, NJ, 2012. The Mathematical Ideas That Animate Great Magic Tricks, With a foreword by Martin Gardner.

[Hal86] Marshall Hall, Jr. *Combinatorial Theory*. Wiley-Interscience Series in Discrete Mathematics. John Wiley & Sons, Inc., New York, second edition, 1986. A Wiley-Interscience Publication.

[LP12] Christian Lawson-Perfect. David’s De Bruijn sequence card trick. Available at [https://aperiodical.com/2012/08/davids-de-bruijn-sequence-card-trick/](https://aperiodical.com/2012/08/davids-de-bruijn-sequence-card-trick/), 2012.

[Son] Siang Wun Song. Magic: Discover 6 in 52 cards knowing their colors based on the De Bruijn sequence. Available at [https://www.ime.usp.br/~song/magic/de-bruijn-magic52cards.pdf](https://www.ime.usp.br/~song/magic/de-bruijn-magic52cards.pdf)
This figure "test.png" is available in "png" format from:

http://arxiv.org/ps/2201.11863v1