Observations on de Bruijn graphs

Tjark Jonathan Garbe

Master's thesis
2020:E44

Lund University

Faculty of Science
Centre for Mathematical Sciences
Mathematics

CENTRUM SCIENTIARUM MATHEMATICARUM

Master's Theses in Mathematical Sciences 2020:E44
ISSN 1404-6342

LUNFMA-3118-2020

Mathematics
Centre for Mathematical Sciences
Lund University
Box 118, SE-221 00 Lund, Sweden

http://www.maths.lth.se/

Abstract of chapter 1

For $n \geq 2$ the number of mixing n-step subshifts of finite type (sft) over the alphabet $\{\mathbf{0}, \mathbf{1}\}$ is proven to be at least $15/16$ times the number of transitive n-step sfts. A conjecture assumes the latter to be at least

$$2^{(3\cdot 2^{n-1}-n)}.$$

Abstract of chapter 2

The *alternating colouring function* is defined. Strings over the alphabet $\{\mathbf{0}, \mathbf{1}\}$ are divided in *colourable* and non-colourable ones. The points in the subshift of finite type defined by forbidding all non-colourable strings of a certain length alternate between states of one colour and states of the other colour. The number of non-colourable strings of length $n \geq 2$ is proven to be $2 \cdot (J_{n-2} + 1)$ where J is the sequence of Jacobsthal numbers.

Contents

Preface

This thesis is to be handed in at Lund University as the written report of the Master's Degree Project of the Master's programme in Mathematics.

Chapter 0 is an introduction to symbolic dynamics and de Bruijn graphs. For the reader with basic knowledge of the notions and notations used in set theory it should provide the definitions and ideas needed to understand the upcoming chapters. The introduction is mainly based on the first four chapters of Lind and Marcus 1995, a book the author wants to recommend for a more detailed understanding.

The book includes the aptly sentence: 'Be warned that terminology in graph theory is notoriously nonstandard.' Different authors use different terminology for the same notions. What this thesis calls vertices and edges other authors call *nodes* and *arcs*, what this thesis calls a walk others call a *path*, what is called a path here others call a *simple path* and so on. Where uncertain the author of this thesis has decided to choose the terminology as it is used by the source that is probably consulted most often in present-day academia: Wikipedia contributors 2020b.

Since Lind and Marcus 1995 do not cover de Bruijn graphs, that information has been taken from Wikipedia contributors 2020a.

Chapter 1 tries to give a lower bound for the number of mixing subshifts of finite type (sft) with a certain step length. To do that first the number of mixing sfts is related to the number of transitive ones. Then a corollary is stated to estimate the number of transitive sfts.

Chapter 2 introduces the *alternating colouring* of strings. Some strings have to be declared non-colourable. Forbidding all those strings naturally leads to an sft.

Appendix A gives the values of the alternating colouring function for strings of length up to and including 7.

To make the main ideas of the thesis accessible to those who paid for the author's education with their tax money (and due to requirement by the university) in appendix B there is a popular scientific summary in English, Swedish and German.

Although chapter 0 focuses on those concepts that will be used in the upcoming chapters, it can be interesting as a general short introduction to graphs and subshifts of finite type. Chapters 1 and 2 only present the results of investigations and are written for the reader intending to build on them. Unless stated, all proofs given in those chapters are due to the author.

Inspired by conventional string indexing chapters, sections, results, equations and figures are numbered starting from zero. To keep the convention in the printed version to have odd pages on the right and even pages on the left page numbering is excepted from that. Section numbers are always given together with the chapter numbers, separated by a dot. Results are numbered inside the sections and their number is separated from the section number by another dot. Equations are also numbered inside the section, but their number is separated from the section number by a hyphen. Also, equation numbers are always given in brackets. Finally figures and tables are numbered directly inside the chapter and their number is separated from the chapter number by a hyphen.

In proofs some equality $=$, implication $\implies$, membership $\in$ and similar relation signs are decorated with an equation reference. If unsure why a certain relation holds the equation referred to should justify it. In the PDF version of this thesis the references are clickable.

Important resources that have not been mentioned so far are the On-Line Encyclopedia of Integer Sequences[1] and the programming language Python, both of which have been very helpful tools. The visualisations of graphs have been drawn using Lucidchart[2]; the text is typeset with LaTeX using Overleaf[3]. The author is grateful for all those who have contributed to LaTeX writing packages or giving advice in the forums of the world wide web. Especially to mention among them is the team behind the KOMA-script which provides the layout for this thesis.

If the popular scientific summary is comprehensible it is thanks to the feedback given by Theresa Hahner. Finally, the author wants to thank his adviser Jörg Schmeling and his examiner Tomas Persson at the Centre for Mathematical Sciences at Lund University for their support throughout the degree project.

Lund, June 2020
Jonathan Garbe

[1] `https://oeis.org/`
[2] `https://www.lucidchart.com/`
[3] `https://www.overleaf.com/`

0 Introduction

0.0 Strings

Let $n \in \mathbb{N}$ and A be a finite set. Then the elements in A^n are called *strings over the alphabet* A. Instead of strings, also the terms *word* and *block* are used. For $w \in A^n$ the *length* n of w is denoted $\#w$. (The same notation is also used for the cardinality of a finite set.) The set A^* is defined as

– *string*

– *length*

– A^*

$$A^* = \bigcup_{n \in \mathbb{N}} A^n.$$

As a shorthand, also notations like the following will be used:

$$A^{\text{even}} = \bigcup_{n \text{ is even}} A^n \qquad\qquad A^{\geq 2} = \bigcup_{n \geq 2} A^n$$

While there are different ways to define A^n (if seeing natural numbers as von Neumann ordinals, one can define it as the set of all functions $n \longrightarrow A$), one has to make sure that for $m \neq n$ the sets A^m and A^n do not intersect for the length of strings over A to be defined. It does not work for example to represent strings of zeroes and ones by their binary representation setting $A = \{0, 1\}$, $A^n = \{0, \ldots, 2^n - 1\}$, because then $0010_2 = 10_2$, so the length is not clear. More relevant than the formal way around this problem is however the notation. The set $\mathcal{A}_2$ is defined as $\{\mathbf{0}, \mathbf{1}\}$ and may be identified with $\{0, 1\}$, but $\mathbf{00}$ is always to be seen as a string of length 2, thereby not equal with $\mathbf{0}$ which is a string of length 1.

– *the set* $\mathcal{A}_2$

For any alphabet A the set A^0 contains exactly one element: The *empty string* ϵ, the only string of length 0. To cover the possible string lengths the natural numbers $\mathbb{N}$ are considered to contain 0.

– *empty string* ϵ

For an alphabet A and two strings $u, v \in A^*$ the *concatenation* uv is a string of length $\#u + \#v$ consisting of the characters of the first string followed by the characters of the second string. For $u = \mathbf{001}, v = \mathbf{0110}$ the concatenation $uv = \mathbf{0010110}$. For $k \in \mathbb{N}$ concatenating u k times with itself is denoted u^k: $u^0 = \epsilon$, $u^4 = uuuu$, $(\mathbf{011})^2 = \mathbf{011011}$ and $u^{k+1} = u^k u$.

– *concatenation*

A string s is called a *substring* of w if there are strings u, v such that $w = usv$. If u can be chosen to be ϵ s is called a *prefix* of w; if v can be ϵ s is a *suffix*. It can be

– *substring*

– *prefix*

– *suffix*

both: **0100** is both a prefix and a suffix of **0100100**. If u, v can at the same time be chosen to be non-empty s is called a *proper infix*. *– proper infix*

The characters in a string are addressed by subscripts:

$$\mathbf{0010110}_0 = \mathbf{0} \qquad \mathbf{0010110}_1 = \mathbf{0} \qquad \mathbf{0010110}_2 = \mathbf{1} \qquad \mathbf{0010110}_6 = \mathbf{0}.$$

The substring of a string w that starts with the character in position $\imath$ and ends with the character in position $\jmath$ where $\imath, \jmath < \#w$ is denoted by $w_{[\imath,\jmath+1)}$. The notation $[\imath, \jmath + 1)$ shall remind of the half open interval including $\imath$ but excluding $\jmath$. The author prefers it because the length of the substring is exactly the difference of the two numbers in the subscript: $\#(\mathbf{0010110})_{[3,5)} = \#(\mathbf{01}) = 2 = 5 - 3$.

Reading a string backwards gives its *reverse*. The reverse of **0010110** is **0110100**. Strings that equal their reverse are called *palindromes*. **0010100** is an example for a palindrome, ϵ another. The notion of a *complement* of a string is restricted to $\mathcal{A}_2^*$: It is the string one gets by substituting the **0**s by **1**s and vice versa. The complement of **0010110** is **1101001**. *– reverse* *– palindrome* *– complement*

The notion of a string can be extended to infinite sequences of characters from a finite alphabet A. An *infinite string* $w \in A^{\mathbb{N}}$ contains exactly one character $w_k \in A$ for each natural number $k \in \mathbb{N}$. $A^{\mathbb{N}}$ is a different object than A^*: The set A^* contains only finite strings, $A^{\mathbb{N}}$ only infinite. *– infinite string*

A *bi-infinite string* $w \in A^{\mathbb{Z}}$ contains a character $w_k \in A$ for each integer $k \in \mathbb{Z}$. The term *full shift* can refer either to $A^{\mathbb{N}}$ or to $A^{\mathbb{Z}}$. In this thesis however shifts are always considered subsets of $A^{\mathbb{N}}$. *– bi-infinite string* *– full shift*

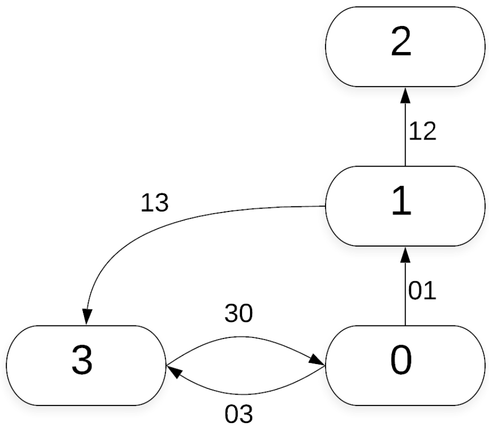

Figure 0-0: The simple digraph $(\{\mathbf{0},\mathbf{1},\mathbf{2},\mathbf{3}\},\{\mathbf{01},\mathbf{03},\mathbf{12},\mathbf{13},\mathbf{30}\})$.

0.1 Graphs

For many objects studied in symbolic dynamics there are different ways to formalise a certain notion. Just as the unit circle can be thought of either as $\mathbb{R}/\mathbb{Z}$ or as a subset of $\mathbb{R}^2$ or $\mathbb{C}$, there are different ways to formally represent an edge, a walk or a subgraph. Each of them has its benefits and usually one will choose the one that is most convenient in a certain situation.

Informally a *directed graph* (short *digraph*) is an object that can be represented by a graphic of points that are connected by arrows. An example gives figure 0-0. The points are called *vertices* or *nodes* while the arrows are called *edges* or *arcs*.

– directed graph
– digraph
– vertex
– edge

If there is at most one edge in each direction between two vertices a digraph is called *simple*, otherwise it is said to be a *multidigraph*.

– simple digraph
– multidigraph

Formally a digraph consists of a vertex set V and an edge set $E \subseteq V^2$ (although even for the edge set another representation will be introduced later). The two elements of an edge are called *starting point* and *endpoint*.

– starting point of an edge
– endpoint of an edge

There are also *undirected graphs* where the edges connect the vertices in both directions instead of being one way roads, however they are not relevant for this thesis. Here a graph should always be understood as a directed one.

Starting from a certain vertex called *starting point*, following a finite number of edges until terminating at an *endpoint* results in a *walk*. In figure 0-0 **13012** is a walk while **0123** is not because there is no edge connecting **2** and **3**. Formally a walk can be defined as a string either in the vertex set (in that case with the condition that any

– starting point of a walk
– endpoint of a walk
– walk

substring of length 2 is contained in the edge set) or in the edge set (in that case with the condition that the starting point of each edge coincides with the endpoint of the previous). The *length* of a walk is the number of edges that are passed which is one less than the number of vertices passed. A walk of length 0 is called *empty*.

– length of a walk
– empty walk

Note a crucial difference between the representations: If defining a walk as a string of edges, there is only one empty walk and that one does not have a defined starting nor endpoint. When seeing a walk as a string of vertices, even an empty walk does have starting and endpoint (which are identical) and hence there are as many walks of length 0 as there are vertices in the graph.

If a walk does not use the same edge twice it is called a *trail*. If it does not even use the same vertex twice, it is called a *path*. Note that any path is a trail as when using the same edge twice one also uses starting and endpoint of that edge twice. Instead of using the string of vertices or edges, a path can also be represented by just the set of edges as the order is clear. If starting and endpoint of a walk coincide, the walk is called *closed*. A closed trail is called a *circuit*. A *cycle* is what comes closest to a closed path: Only start and endpoint of a cycle coincide; all other vertices are visited only once.

– trail
– path
– closed walk
– circuit
– cycle

This thesis will usually use the word cycle for a set of edges: In that case, start and endpoint are not determined. A cycle passes equally many edges as vertices and the length of it is at most the length of the vertex set. For some digraphs there are cycles that actually pass every single vertex. Such cycles are called *Hamiltonian*.

– Hamiltonian cycle

Infinite walks and *bi-infinite walks* are natural extensions of the concept of a walk. They are (bi-) infinite strings such that their substrings of length 2 are edges in the digraph. The set of all (bi-) infinite walks in a certain digraph is called the *vertex shift* of that digraph.

– infinite walk
– bi-infinite walk
– vertex shift

A vertex that is not the endpoint for any edge is called a *source*. Walks that do not start at a source can never reach it. The poetic metaphor of flowing water is also used for naming the counterpart: A vertex that is not the starting point for any edge is called a *sink*. Walks cannot start at sinks but they may end there. As an umbrella term, sources and sinks are said to be *stranded*. A vertex that is both a source and a sink is called *isolated*.

– source
– sink
– stranded vertex
– isolated vertex

A digraph is said to be *essential* if it does not contain any stranded vertices. It is *strongly connected* if there is a walk from any vertex to any other vertex. Every strongly connected digraph is essential, but not vice versa.

– essential digraph
– strongly connected digraph

Here a difference between $A^{\mathbb{N}}$ and $A^{\mathbb{Z}}$ becomes apparent: While sinks can never appear in infinite nor in bi-infinite walks, infinite walks can start at a source while bi-infinite walks cannot contain any stranded vertices at all.

A digraph (V', E') is said to be a *subgraph* of another digraph (V, E) if $V' \subseteq V$ and $E' \subseteq E$. Note that one cannot just choose any subsets because all starting and endpoints of the edges in E' must be included in V'. (V', E') is *spanning* if $V' = V$.

– subgraph
– spanning subgraph

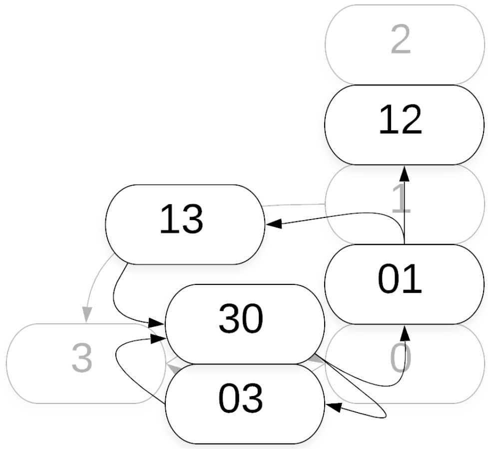

Figure 0-1: The line graph of the graph in figure 0-0. The original graph can be seen faded in the background. The new vertices are what previously were the edges. The new edges correspond two the walks of length 2 in the original graph.

The *line graph* (E, L) of a digraph (V, E) is a digraph constructed by setting – *line graph*

$$L = \Big\{\big((u, v), (w, x)\big) \in E^2; v = w\Big\}.$$

The new vertices of the line graph are the old edges and two of them are connected by a new edge if and only if the endpoint of one old edge is the starting point of the next. Figure 0-1 shows the line graph of the digraph in figure 0-0.

In his proof de Bruijn 1946 already uses the notion of a line graph but calls it *doubling*.

The walks in a digraph represented by strings of edges are exactly the walks in the line graph represented by strings of vertices. However, the original walk is one step longer than the associated walk in the line graph.

Following this definition $L \subseteq (V^2)^2 \cong V^4$. However, edges could also be seen as elements in V^3: Instead of

$$\{(\mathbf{01}, \mathbf{12}), (\mathbf{01}, \mathbf{13}), (\mathbf{03}, \mathbf{30}), (\mathbf{13}, \mathbf{30}), (\mathbf{30}, \mathbf{01}), (\mathbf{30}, \mathbf{03})\}$$

in figure 0-1 the edge set can shortly be expressed as

$$\{\mathbf{012}, \mathbf{013}, \mathbf{030}, \mathbf{130}, \mathbf{301}, \mathbf{303}\}.$$

$$\begin{array}{c} \\ \mathbf{0} \\ \mathbf{1} \\ \mathbf{2} \\ \mathbf{3} \end{array} \begin{array}{c} \begin{array}{cccc} \mathbf{0} & \mathbf{1} & \mathbf{2} & \mathbf{3} \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \end{array} \qquad \begin{array}{c} \\ \mathbf{01} \\ \mathbf{03} \\ \mathbf{12} \\ \mathbf{13} \\ \mathbf{30} \end{array} \begin{array}{c} \begin{array}{ccccc} \mathbf{01} & \mathbf{03} & \mathbf{12} & \mathbf{13} & \mathbf{30} \end{array} \\ \begin{pmatrix} 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

Figure 0-2: The adjacency matrix for the digraph in figure 0-0 and the one for its line graph which is shown in figure 0-1.

$$\begin{array}{c} \\ \mathbf{0} \\ \mathbf{1} \\ \mathbf{2} \\ \mathbf{3} \end{array} \begin{array}{c} \begin{array}{cccc} \mathbf{0} & \mathbf{1} & \mathbf{2} & \mathbf{3} \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \end{array} \qquad \begin{array}{c} \\ \mathbf{0} \\ \mathbf{1} \\ \mathbf{2} \\ \mathbf{3} \end{array} \begin{array}{c} \begin{array}{cccc} \mathbf{0} & \mathbf{1} & \mathbf{2} & \mathbf{3} \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 \end{pmatrix} \end{array} \qquad \begin{array}{c} \\ \mathbf{0} \\ \mathbf{1} \\ \mathbf{2} \\ \mathbf{3} \end{array} \begin{array}{c} \begin{array}{cccc} \mathbf{0} & \mathbf{1} & \mathbf{2} & \mathbf{3} \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 2 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix} \end{array}$$

Figure 0-3: The 2^{nd}, 3^{rd} and 4^{th} power of the matrix on the left of figure 0-2. The graph corresponding to the 2^{nd} power is shown in figure 0-4.

The vertex shift of the line graph is called the *edge shift* of the original graph.

– edge shift

Another representation of a digraph with vertex set V is its *adjacency matrix* $m \in \{0,1\}^{V\times V}$. For $u, v \in V$ the entry $m_{u,v}$ is 1 if there is an edge connecting u and v and 0 otherwise. The adjacency matrix of the graph in figure 0-0 is given in figure 0-2. Note that the row **2** contains only zeroes. That corresponds to the vertex **2** being a sink. A source would correspond to a column of zeroes. A matrix having at least a one in each row and column hence corresponds to an essential graph and is therefore itself called *essential*.

– adjacency matrix

– essential matrix

For a multidigraph the adjacency matrix is found in $\mathbb{N}^{V\times V}$ rather than $\{0,1\}^{V\times V}$. The entry $m_{u,v} \in \mathbb{N}$ then gives information on how many edges there are connecting u and v.

A perk of adjacency matrices is the information their powers provide: For $L \in \mathbb{N}$ the matrix entry $(m^L)_{u,v}$ is the number of walks of length L there are from u to v. The multidigraph that corresponds to the L^{th} power of the adjacency matrix of a digraph is also called the L^{th} *power* of that digraph. The 2^{nd} power of the digraph is

– power of a digraph

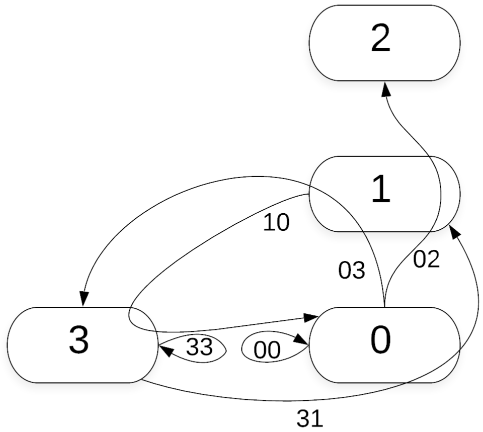

Figure 0-4: The 2^{nd} power graph of the digraph in figure 0-0. The vertex set is the same as in the original digraph. With the exception of **00** and **33** the edges are drawn such that they touch the vertex the corresponding walk of length 2 passed in the original graph.

---

shown in figure 0-4, its adjacency matrix together with those of even higher power in figure 0-3. The 2^{nd} power of a graph and its line graph are related: They have equally many edges because both in the 2^{nd} power graph and in the line graph each edge correspond to a walk of length 2 in the original graph. However a crucial difference is that powers preserve the vertex set; line graphs do not.

The question whether there is a walk from an edge $u \in V$ to another edge $v \in V$ can be reformulated in terms of the adjacency matrix m by asking whether there is an $L \in \mathbb{N}$ such that $(m^L)_{u,v} > 0$. If for any $u, v \in V$ there is such an L the matrix is called *irreducible*. Therefore a digraph is strongly connected if and only if its adjacency matrix is irreducible. *– irreducible matrix*

Since L may depend on u and v even if a matrix m is irreducible that does not mean there is an $L \in \mathbb{N}$ such that every entry in m^L is positive. If there is an L such that m^L contains only positive values, m is called *primitive*. Since a primitive matrix m must be irreducible and an irreducible matrix essential, m contains no row or column of zeroes, so in fact for any $l \geq L$ the matrix m^l contains only positive entries. *– primitive matrix*

For any vertex set V, the digraph (V, V^2) where there is an edge in both directions between any two vertices is called *complete*. The complete digraph is the line graph of a multidigraph with only one vertex but $\#V$ edges from that vertex to itself. Any digraph is a subgraph of the complete digraph of the vertex set. The adjacency matrix *– complete digraph*

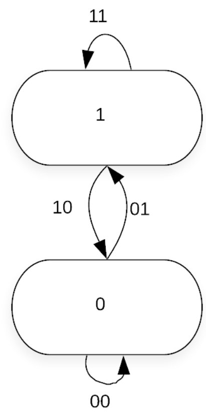

Figure 0-5: The 1-dimensional de Bruijn graph over the alphabet $\mathcal{A}_2$. It is the complete graph over $\mathcal{A}_2$.

of a complete digraph is the matrix filled with ones.

0.2 Sfts

The complete digraph with vertex set A is also called the 1*-dimensional de Bruijn graph* over the alphabet A. For $n \geq 1$ the *n-dimensional de Bruijn graph* is the line graph of the $(n-1)$-dimensional de Bruijn graph. Figures 0-5, 0-6, 0-7 and 0-8 show the 1-, 2-, 3- and 4-dimensional de Bruijn graphs over the alphabet $\mathcal{A}_2$, figure 0-9 their adjacency matrices.

– de Bruijn graph

Here the idea just mentioned has been extended: Instead of seeing the edge set of the 3-dimensional de Bruijn graph as a subset of $((\mathcal{A}_2^2)^2)^2$, they are just seen as a subset of $\mathcal{A}_2^4$. In fact the edge set is equivalent to $\mathcal{A}_2^4$ and will from now on be treated as being equal to $\mathcal{A}_2^4$. In general, the edge set of the n-dimensional de Bruijn graph is $\mathcal{A}_2^{n+1}$ while the vertex set is $\mathcal{A}_2^n$. Two vertices $u, v \in A^n$ are connected by an edge if and only if the suffix of length $n-1$ of u equals the prefix of length $n-1$ of v, meaning that u and v have a large intersection. The starting point of an edge is its prefix of length n, the endpoint is the suffix of length n.

For $n \in \mathbb{N}$ a walk in the n-dimensional de Bruijn graph over the alphabet A would be a string over the alphabet A^n (an element in $(A^n)^*$ that is). An example for a path in the 3-dimensional de Bruijn graph over $\mathcal{A}_2$ would be $(\mathbf{001}, \mathbf{010}, \mathbf{101}, \mathbf{011}, \mathbf{110})$. The path is shown in figure 0-10. However a natural projection keeping the starting

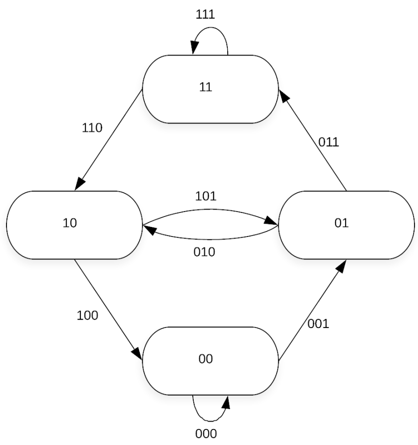

Figure 0-6: The 2-dimensional de Bruijn graph over the alphabet $\mathcal{A}_2$. It is the line graph of the 1-dimensional de Bruijn graph.

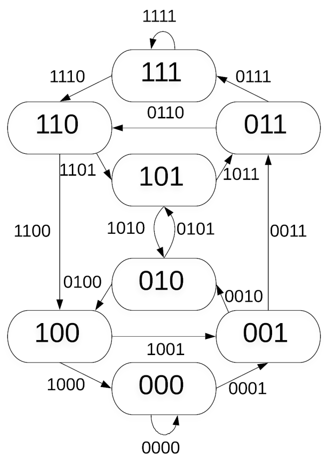

Figure 0-7: The 3-dimensional de Bruijn graph over the alphabet $\mathcal{A}_2$

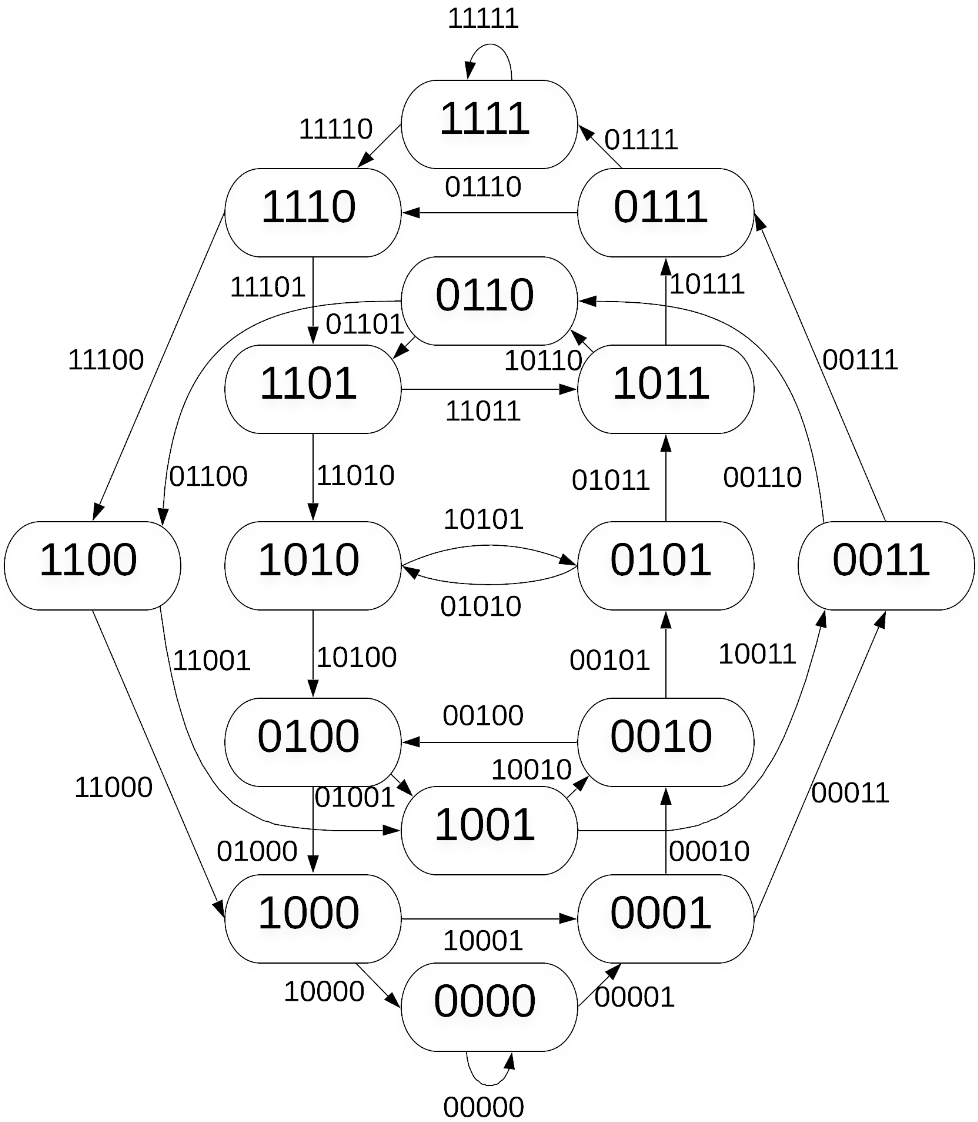

Figure 0-8: The 4-dimensional de Bruijn graph over the alphabet $\mathcal{A}_2$

$$\begin{array}{c} \\ \mathbf{0} \\ \mathbf{1} \end{array}\begin{array}{c} \begin{array}{cc} \mathbf{0} & \mathbf{1} \end{array} \\ \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \end{array} \qquad \begin{array}{c} \\ \mathbf{00} \\ \mathbf{01} \\ \mathbf{10} \\ \mathbf{11} \end{array}\begin{array}{c} \begin{array}{cccc} \mathbf{00} & \mathbf{01} & \mathbf{10} & \mathbf{11} \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \end{array}$$

$$\begin{array}{c} \\ \mathbf{000} \\ \mathbf{001} \\ \mathbf{010} \\ \mathbf{011} \\ \mathbf{100} \\ \mathbf{101} \\ \mathbf{110} \\ \mathbf{111} \end{array}\begin{array}{c} \begin{array}{cccccccc} \mathbf{000} & \mathbf{001} & \mathbf{010} & \mathbf{011} & \mathbf{100} & \mathbf{101} & \mathbf{110} & \mathbf{111} \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \end{array}$$

Figure 0-9: The adjacency matrices of the 1-, 2- and 3-dimensional de Bruijn graph over the alphabet $\mathcal{A}_2$

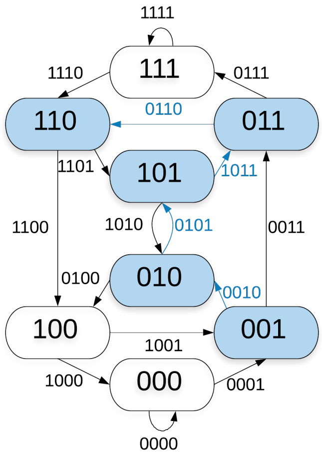

Figure 0-10: The 3-dimensional de Bruijn graph. The vertices and edges that are part of the path **0010110** are highlighted blue. This shows why it is sufficient to give a set of edges to define a path: The vertices and their order are then clear. **001** is the starting point of the path; **110** is its endpoint.

---

point completely but only the last character from the following vertices preserves the information while collapsing the expression to **0010110**.

In general, walks in the n-dimensional de Bruijn graph will be seen as elements in $\mathcal{A}_2^{\geq n}$. If $k, n \geq 1$ and $w \in A^{k+n}$ shall be considered a walk in the n-dimensional de Bruijn graph, $w_{[0,n)}$ is the starting point, $w_{[0,n+1)}$ is the first edge, $w_{[1,1+n)}$ is the second vertex the walk passes, $w_{[k,k+n)}$ is the endpoint.

One has to be careful that the length of the string differs from the length of the walk it represents. As a walk w has length k because it contains k substrings of length $n+1$ representing an edge each. The length of w as a string however is $k+n$. $\#w$ shall always refer to the latter.

For $n \in \mathbb{N}$ an *n-dimensional de Bruijn subgraph* is a spanning subgraph of the n-dimensional de Bruijn graph. $w \in A^{\geq n}$ is a walk in a certain de Bruijn subgraph if and only if all substrings of length $n+1$ are in the edge set. – *de Bruijn subgraph*

Using this shorthand to express walks, also the notion of infinite walks has to be refined accordingly: For $n \in \mathbb{N}$ the infinite walks in an n-dimensional de Bruijn subgraph are the infinite strings whose substrings of length $n+1$ can be found in the edge set. The vertex shift one thus gets is called an *n-step shift of finite type*, abbreviated *sft*. The edge shift of an n-dimensional de Bruijn graph equals the vertex shift of the $(n+1)$-dimensional de Bruijn graph. – *n-step shift of finite type* – *sft*

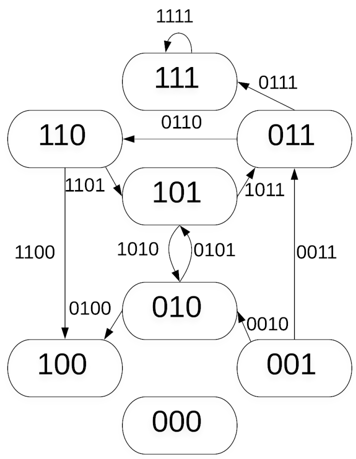

Figure 0-11: A 3-dimensional de Bruijn subgraph. Here **000**, **001**, **100** are stranded, **000**, **001** are sources, **000**, **100** are sinks, **000** is isolated. **111** is not a sink and can occur in the associated sft. The graph is not strongly connected, so the sft is not transitive.

Alternatively, an sft can be constructed by taking a finite set F of finite strings and removing from the full shift any point that contains an element in F. One might ask the question what happens if F contains strings of different length. However, instead of forbidding a string w of length 3 one can equivalently forbid all strings of length 5 that contain w. This observation points to a problem: While different sfts must arise from different de Bruijn subgraphs, two different de Bruijn subgraphs can lead to the same sft. Any de Bruijn subgraph not containing a closed walk for example will lead to an empty shift. An important exception will be given in lemma 1.0.0.

There is also a more general notion of a *shift* that is constructed by forbidding an arbitrary (and possibly infinite) set of substrings. – *shift*

A shift becomes a dynamic object by introducing the *shift operator* $\sigma : A^{\mathbb{Z}} \to A^{\mathbb{Z}}$ for some alphabet A as – *shift operator*

$$\big(\sigma(w)\big)_k = w_{k+1}.$$

Note that an sft (and in fact any shift) is closed under the shift operator.

When looking at a certain string one can say that with each time step that passes, the string is shifted one place to the left. Alternatively one can focus on the place with index 0 and consider that as the current state. With each time step passing a new character appears. Instead of imagining an object that spreads infinitely long in space one can think of a single character that somehow also contains information

Graph		Adjacency matrix		Vertex shift
essential	$\Longleftrightarrow$	*essential*		
$\Uparrow$		$\Uparrow$		
strongly connected	$\Longleftrightarrow$	*irreducible*	$\Longleftrightarrow$	*transitive*
$\Uparrow$		$\Uparrow$		$\Uparrow$
primitive	$\Longleftrightarrow$	*primitive*	$\Longleftrightarrow$	*mixing*

Table 0-12: Different names for equivalent properties, expressed in terms of graphs, adjacency matrices and vertex shifts.

---

about how it is going to change over time. Say, a machine with an 7-segment display showing a new character each second.

Now another difference between $A^{\mathbb{N}}$ and $A^{\mathbb{Z}}$ can be formulated: In the latter object the shift operator is invertible, the shift is a reversible system. Restricting σ to $A^{\mathbb{N}}$ leads to a loss of information in each step.

– *transitive sft*

The dynamical property of an sft being *transitive* corresponds to the de Bruijn subgraph being strongly connected. It means that whenever both u and v occur as a substring somewhere in a certain sft, then there is also a single point containing first u and at some later position v.

– *mixing sft*

An sft is *mixing* if its de Bruijn subgraph is primitive. Table 0-12 relates different properties of graphs, their adjacency matrices and vertex shifts. In the literature, the terms are sometimes used interchangeably, also graphs and sfts can for example be called irreducible.

1 Counting the mixing sfts

1.0 Counting the mixing sfts among the transitive ones

Lemma 1.0.0. *For $n \in \mathbb{N}$ the number of transitive n-step sfts equals the number of strongly connected n-dimensional de Bruijn subgraphs.*

Proof. Each transitive n-step sft is described by an n-dimensional de Bruijn subgraph, so there are at least as many n-dimensional de Bruijn subgraphs as there are transitive n-step sfts. It is left to show that different strongly connected de Bruijn subgraphs of the same dimension describe different sfts.

If two strongly connected de Bruijn subgraphs of the same dimensions are not equal, one of them contains an edge the other does not. Since the digraph which contains that edge is strongly connected there is a closed walk passing that edge and so there is an infinite walk doing so. That infinite walk is a point in one sft but not in the other so the sfts differ.

Corollary 1.0.1. *For $n \in \mathbb{N}$ the number of mixing n-step sfts equals the number of primitive n-dimensional de Bruijn subgraphs.*

Proof. Since an sft is mixing if and only if it is described by a primitive de Bruijn subgraph and because a mixing sft is transitive the statement follows from lemma 1.0.0.

Lemma 1.0.2. *Let $n, p, q, r, s \in \mathbb{N}$ such that*

$$n \geq s \cdot q^2 \wedge r \cdot p - s \cdot q = 1. \tag{1.0-0}$$

Then

$$\exists t, u \in \mathbb{N} \quad t \cdot p + u \cdot q = n.$$

Proof. The statement is trivial for $q = 0$, so let $q > 0$. Pick $l, m \in \mathbb{N}$ such that

$$n = l \cdot q + m \wedge m < q. \tag{1.0-1}$$

Then

$$l \cdot q \overset{(1.0\text{-}1)}{=} n - m \overset{(1.0\text{-}1)}{>} n - q \overset{(1.0\text{-}0)}{\geq} s \cdot q^2 - q = \left(s \cdot q - 1\right) \cdot q,$$

so $l \geq s \cdot q \geq s \cdot m$. Set $t = m \cdot r, u = l - s \cdot m$. Then

$$\begin{aligned} t \cdot p + u \cdot q &= m \cdot r \cdot p + l \cdot q - s \cdot m \cdot q \\ &= m \cdot \left(r \cdot p - s \cdot q\right) + l \cdot q \\ &\overset{(1.0\text{-}0)}{=} m + l \cdot q \\ &\overset{(1.0\text{-}1)}{=} n. \end{aligned}$$

Corollary 1.0.3. *Let a strongly connected digraph contain two closed walks with lengths p and q respectively such that $r \cdot p - s \cdot q = 1$. Let m be the number of steps needed to go from any vertex to any other vertex and*

$$l \geq 3 \cdot m + s \cdot q^2. \tag{1.0-2}$$

Then for any vertices X, Y there is a walk of length l from X to Y.

Proof. Let
P be a vertex in the closed walk of length p,
Q a vertex in the closed walk of length q,
$\imath \leq m$ the length of a walk from X to P,
$\jmath \leq m$ the length of a walk from P to Q,
$k \leq m$ the length of a walk from Q to Y.
Set

$$n = l - \left(\imath + \jmath + k\right). \tag{1.0-3}$$

Then

$$n \overset{(1.0\text{-}3)(1.0\text{-}2)}{\geq} 3 \cdot m + s \cdot q^2 - 3 \cdot m = s \cdot q^2,$$

so by lemma 1.0.2 there are $t, u \in \mathbb{N}$ such that

$$t \cdot p + u \cdot q = n. \tag{1.0-4}$$

Consider the following walk:

- Go from X to P,
- follow t times the closed walk of length p,
- go to Q,

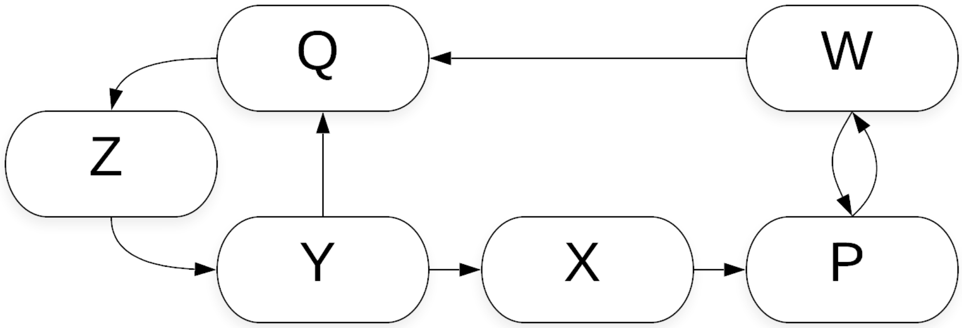

Figure 1-0: An example for a strongly connected digraph with two closed walks whose lengths are relatively prime, in this case 2 and 3

- follow u times the closed walk of length q,
- go to Y.

The length of the walk is

$$\imath + t \cdot p + \jmath + u \cdot q + k \overset{(1.0\text{-}4)}{=} n + \imath + \jmath + k \overset{(1.0\text{-}3)}{=} l.$$

Example. Consider the graph in figure 1-0. It is strongly connected; any vertex can be reached from another by a walk of at most length 5. Moreover it contains a cycle of length 2 and another one of length 3. Using the letters in which corollary 1.0.3 is formulated one can set $p = 2, q = 3, r = 2, s = 1, m = 5$. Hence for any number $l \geq 3 \cdot m + s \cdot q^2 = 24$ there should be a walk of length l from vertex X to vertex Y. For example set $l = 28$. The n used in the proof becomes 23 and a possible choice for t and u is $t = 10, u = 1$. That leads to the walk $XP(WP)^{10}WQ(ZYQ)^{1}ZY$ which can be shortened to $X(PW)^{11}(QZY)^2$ and passes $1 + 11 \cdot 2 + 2 \cdot 3 = 29$ vertices so has the desired length 28.

Corollary 1.0.4. *A strongly connected digraph is primitive if and only if it contains two closed walks whose lengths are relatively prime.*

Proof. Recall that a digraph is called primitive if there is an $L \in \mathbb{N}$ such that for any $l \geq L$ there is a walk of length l from any vertex to any other vertex. The existence of such a walk under the conditions given has been proven in corollary 1.0.3.

To see the reverse implication note that any primitive digraph contains closed walks of length L and of length $L + 1$ which are relatively prime.

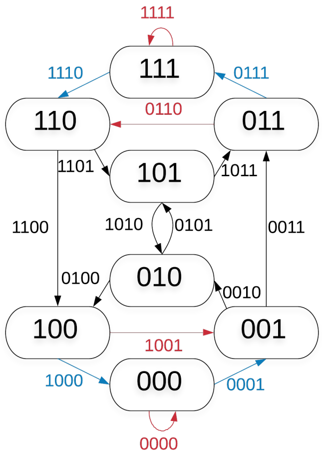

Figure 1-1: The 3-dimensional de Bruijn graph. The edges in Λ^4 are coloured blue. Removing any of them would disconnect the graph. The edges in Ξ^4 are coloured red. As long as the edges in Λ^4 are present, they can be removed without disconnecting the graph.

Remark. Corollary 1.0.4 is a formulation of theorem 4.5.8 in Lind and Marcus 1995. There it is proven using the adjacency matrix rather than the graph itself.

Notation. Let $n \geq 2$. The following notation will be used:

$$\Xi^n = \left\{\mathbf{0}^n, \mathbf{0}\mathbf{1}^{n-2}\mathbf{0}, \mathbf{1}\mathbf{0}^{n-2}\mathbf{1}, \mathbf{1}^n\right\}$$
$$\Lambda^n = \left\{\mathbf{0}^{n-1}\mathbf{1}, \mathbf{0}\mathbf{1}^{n-1}, \mathbf{1}\mathbf{0}^{n-1}, \mathbf{1}^{n-1}\mathbf{0}\right\}$$

Lemma 1.0.5. *Let $n \geq 1$ and $c \subseteq \mathcal{A}_2^{n+1}$ be a strongly connected n-dimensional de Bruijn subgraph. Then*

$$\Lambda^{n+1} \subseteq c.$$

Proof. $\mathbf{0}^n\mathbf{1}$ and $\mathbf{1}^n\mathbf{0}$ are the only edges leading away from $\mathbf{0}^n$ and $\mathbf{1}^n$ respectively while $\mathbf{1}\mathbf{0}^n$ and $\mathbf{0}\mathbf{1}^n$ are the only ones leading to them. Dropping any of them would disconnect the graph.

Lemma 1.0.6. *Let $n \geq 1$ and $c \subseteq \mathcal{A}_2^{n+1}$ be a strongly connected n-dimensional de Bruijn subgraph. Then $c \setminus \Xi^{n+1}$ is still strongly connected.*

Proof. Consider $w \in \mathcal{A}_2^{\geq n}$ as a walk in c (meaning that $\forall k < \#w - n \quad w_{[k,k+n+1)} \in c$). The substrings $\mathbf{0}^{n+1}, \mathbf{1}^{n+1}$ of w can be replaced by $\mathbf{0}^n, \mathbf{1}^n$ which are necessarily allowed in w due to the fact that c is spanning. The substrings $\mathbf{01}^{n-1}\mathbf{0}, \mathbf{10}^{n-1}\mathbf{1}$ of w can be replaced by $\mathbf{01}^n\mathbf{0}, \mathbf{10}^n\mathbf{1}$ which are allowed according to lemma 1.0.5.

Corollary 1.0.7. *Let $n \geq 1$ and $c \subseteq \mathcal{A}_2^{n+1}$ be a strongly connected n-dimensional de Bruijn subgraph such that $c \cap \Xi^{n+1} \neq \emptyset$. Then c is primitive.*

Proof. Pick a closed walk $w \in \mathcal{A}_2^{\geq n}$ in c that passes all edges in Λ^{n+1}. Create another walk $\tilde{w} \in \mathcal{A}_2^{\geq n}$ in c by performing exactly one of the substitutions described in the proof of lemma 1.0.6, in one or the other direction. Then $|\#\tilde{w} - \#w| = 1$, so $\#\tilde{w}$ and $\#w$ are relatively prime. Hence by corollary 1.0.4 c is primitive.

Counterexample. The condition given in corollary 1.0.7 is sufficient, but not necessary. Consider $\mathcal{A}_2^5$ as the edge set of the 4-dimensional de Bruijn graph. $\mathcal{A}_2^5 \setminus \Xi^5$ is still mixing.

Corollary 1.0.8. *Let $n \geq 2$ and C_n be the set of all strongly connected n-dimensional de Bruijn subgraphs. Then $\#C_n$ is divisible by* 16.

Proof. Set $\tilde{C}_n = \{c \setminus \Xi^{n+1}; c \in C_n\}$. According to lemma 1.0.6 still $\tilde{C}_n \subseteq C_n$, so

$$C_n = \left\{a \cup c; a \subseteq \Xi^{n+1} \wedge c \in \tilde{C}_n\right\}$$

and hence $\#C_n = \#(\mathcal{P}\Xi^{n+1} \times \tilde{C}_n) = 16 \cdot \tilde{C}_n$.

Theorem 1.0.9. *Let $n \geq 2$, C_n be the set of all strongly connected n-dimensional de Bruijn subgraphs and $M_n \subseteq C_n$ the primitive ones among them. Then*

$$\#M_n \geq \frac{15}{16} \cdot \#C_n. \tag{1.0-5}$$

Proof. Set

$$\tilde{M}_n = \left\{a \cup c; a \subseteq \Xi^{n+1} \wedge a \neq \emptyset \wedge c \in C_n \setminus M_n\right\}. \tag{1.0-6}$$

According to corollary 1.0.7 $\tilde{M}_n \subseteq M_n$, so

$$\begin{aligned}
\#M_n \geq \#\tilde{M}_n &\overset{(1.0\text{-}6)}{=} \#\Big(\big(\mathcal{P}\Xi^{n+1} \setminus \{\emptyset\}\big) \times \big(C_n \setminus M_n\big)\Big) \\
&= \Big(\#\big(\mathcal{P}\Xi^{n+1}\big) - 1\Big) \cdot \big(\#C_n - \#M_n\big) \\
&= 15 \cdot \big(\#C_n - \#M_n\big).
\end{aligned}$$

The statement follows from some linear manipulations.

n	1	2	3	4
$\#C_n$	4	16	560	1215712

Table 1-2: The number $\#C_n$ of strongly connected n-dimensional de Bruijn subgraphs. For $n \in \{3,4\}$ the numbers have been computed with Python and the networkx package.

Corollary 1.0.10. *For $n \geq 2$ the number of mixing n-step sfts is at least $15/16$ times the number of transitive ones.*

Proof. This follows directly from theorem 1.0.9 and lemma 1.0.0 with corollary 1.0.1.

Counterexample. Both equality and strict inequality occur:

$$\#M_2 = \frac{15}{16} \cdot \#C_2$$
$$\#M_4 > \frac{15}{16} \cdot \#C_4$$

1.1 Counting the transitive sfts

Notation. Let a be a path through a simple digraph (thought of as a set of edges) and Q a vertex that a passes. Then $a(Q)$ denotes the vertex such that $(Q, a(Q)) \in a$. Similarly, $a^{-1}(Q)$ is defined such that $(a^{-1}(Q), Q) \in a$.

Lemma 1.1.0. *Let (V, E) be a simple digraph including two different but intersecting Hamiltonian cycles. Let C be the set of strongly connected, spanning subgraphs of (V, E). Then*

$$\#C \geq 2 \cdot 2^{\#E - \#V}.$$

Proof. Let $a, b \subseteq E$ be Hamiltonian cycles such that $a \cap b \neq \emptyset$. Set

$$A = \{c \subseteq E; a \subseteq c\}$$
$$B = \left\{c \subseteq E; b \subseteq c\right\}.$$

Then

$$\#A = \#\left\{a \cup d; d \subseteq E \setminus a\right\} = \#\Big(\mathcal{P}\big(E \setminus a\big)\Big) = 2^{\#E - \#a} = 2^{\#E - \#V} = \#B,$$

so

$$\begin{aligned}\#C &= \#\big(A \cup B\big) + \#\Big(C \setminus \big(A \cup B\big)\Big) \\ &= \#A + \#B - \#\big(A \cap B\big) + \#\Big(C \setminus \big(A \cup B\big)\Big) \\ &= 2 \cdot 2^{\#E - \#V} - \#\big(A \cap B\big) + \#\Big(C \setminus \big(A \cup B\big)\Big).\end{aligned}$$

The aim is now to construct an injection $f : A \cap B \rightarrowtail C \setminus (A \cup B)$. Note that $a \cap b$ cannot be a closed path. (In fact, usually it will not even be a path.) Since (V, E) is simple, there are vertices Q, R such that there is a path from Q to R in $a \cap b$ but $a^{-1}(Q) \neq b^{-1}(Q) \wedge a(R) \neq b(R)$, meaning that Q, R denote the starting and endpoint of a locally maximal path in $a \cap b$. For $c \in A \cap B$ set

$$f(c) = c \setminus \Big\{ \Big(a^{-1}(Q), Q\Big), \big(R, b(R)\big) \Big\}.$$

Since

$$\Big\{ \Big(a^{-1}(Q), Q\Big), \big(R, b(R)\big) \Big\} \subseteq c$$

f is injective and clearly $f(c) \notin A \cup B$. To see that $f(c)$ is connected note that whenever there is no path from a certain vertex to another in $f(c) \cap a$, there is a path in $f(c) \cap b$.

Remark. Here the Hamiltonian cycles are regarded as sets of edges. Besides that being convenient to formulate the proof it is important that two cycles are regarded as different only if they actually pass different edges. Counting the same cycle starting at different vertices several times would lead to wrong numbers.

Counterexample. The inequality in lemma 1.1.0 cannot be replaced by equality: In the 3-dimensional de Bruijn graph there are two intersecting Hamiltonian cycles and $\#V = 8, \#E = 16$. However

$$\#C = 560 > 512 = 2 \cdot 2^{\#E - \#V}.$$

Fact 1.1.1. *Let $n \geq 1$ and $\#H_n$ be the number of Hamiltonian cycles in the n-dimensional de Bruijn graph. Then*

$$\#H_n = 2^{(2^{n-1} - n)}. \qquad (1.1\text{-}0)$$

n	1	2	3	4	5	6
$\#H_n$	1	1	2	16	2048	67108864

Table 1-3: The first six members of de Bruijn's sequence. $\#H_n$ is the number of Hamiltonian cycles in the n-dimensional de Bruijn graph.

Remark. Fact 1.1.1 was famously proven by de Bruijn 1946, after whom $\#H$ was called the *de Bruijn sequence* and which also led to the name de Bruijn graph. However, de Bruijn 1975 discovered that Flye Sainte-Marie 1894 already had published a proof of the same fact.

– de Bruijn sequence

Conjecture 1.1.2. *Let H be a set of Hamiltonian cycles in (V, E) such that $\bigcap H \neq \emptyset$. Let C be the number of strongly connected, spanning subgraphs of (V, E). Then*

$$\#C \geq \#H \cdot 2^{\#E - \#V}. \tag{1.1-1}$$

Remark. For $\#H \leq 1$ the conjecture holds trivially while the case $\#H = 2$ has been established as lemma 1.1.0.

Corollary 1.1.3. *Let $n \geq 1$ and C_n be the set of strongly connected n-dimensional de Bruijn subgraphs. Then*

$$\#C_n \geq 2^{(3 \cdot 2^{n-1} - n)}. \tag{1.1-2}$$

Proof. Let $H_n \subseteq \mathcal{A}_2^{n+1}$ be the set of Hamiltonian cycles in the n-dimensional de Bruijn graph. According to lemma 1.0.5 $\bigcap H_n \supseteq \Lambda^{n+1} \neq \emptyset$, so

$$\#C_n \overset{(1.1\text{-}1)}{\geq} \#H_n \cdot 2^{(2^{n+1} - 2^n)} \overset{(1.1\text{-}0)}{=} 2^{(2^{n-1} - n)} \cdot 2^{(2^{n+1} - 2^n)} = 2^{(3 \cdot 2^{n-1} - n)}.$$

Corollary 1.1.4. *Let $n \geq 2$ and M_n be the number of primitive n-dimensional de Bruijn subgraphs. Then*

$$\#M_n \geq 15 \cdot 2^{(3 \cdot 2^{n-1} - n - 4)}.$$

Proof.

$$\#M_n \overset{(1.0\text{-}5)}{\geq} \frac{15}{16} \cdot \#C_n \overset{(1.1\text{-}2)}{\geq} \frac{15}{16} \cdot 2^{(3 \cdot 2^{n-1} - n)} = 15 \cdot 2^{(3 \cdot 2^{n-1} - n - 4)}$$

Remark. The proofs of corollaries 1.1.3 and 1.1.4 assume conjecture 1.1.2 to hold.

Remark. Always have lemma 1.0.0 and corollary 1.0.1 in mind connecting the number of transitive and mixing sfts with the number of strongly connected and primitive de Bruijn subgraphs, respectively.

2 Alternating Colouring

The aim of this chapter is to define the alternating colouring function ψ which has some interesting properties. Figures 2-10 and 2-13 give an impression of how ψ structures de Bruijn graphs and infinite strings. The definition will however not be given earlier than in the beginning of section 2.4 because it uses the functions ξ and ϕ which shall be introduced and studied first.

2.0 Notation

Notation. Define $l, r : \mathcal{A}_2^{\geq 1} \to \mathcal{A}_2^*, m : \mathcal{A}_2^{\geq 2} \to \mathcal{A}_2^*, R, C : \mathcal{A}_2^* \to \mathcal{A}_2^*, T, CT : \mathbb{N} \to \mathcal{A}_2^*$ by

– l, r, m, R, C, T, CT

$$\begin{aligned} l(w) &= w_{[0,\#w-1)} \\ r(w) &= w_{[1,\#w)} \\ m &= l \circ r \\ R(w_0 \ldots w_{\#w-1}) &= w_{\#w-1} \ldots w_0 \\ C(w_0 \ldots w_{\#w-1}) &= C(w_0) \ldots C(w_{\#w-1}), \text{ where } C(\mathbf{0}) = \mathbf{1} \wedge C(\mathbf{1}) = \mathbf{0} \\ T^n &= \begin{cases} \epsilon & \text{if } n = 0 \\ T^{n-1}\mathbf{0} & \text{if } n \text{ odd} \\ T^{n-1}\mathbf{1} & \text{if } n \text{ even} > 0 \end{cases} \\ CT^n &= C(T^n). \end{aligned}$$

Remark. $l(w), r(w)$ are prefix and suffix with length $\#w - 1$ of w. $m(w)$ is the proper infix of length $\#w - 2$. $R(w)$ is the reverse of w (meaning that w is a palindrome if and only if $R(w) = w$) while $C(w)$ is its complement. T^n, CT^n are the strings of length n starting with $\mathbf{0}, \mathbf{1}$ respectively and alternating between $\mathbf{0}$ and $\mathbf{1}$.

Example.

$$\begin{aligned} l(\mathbf{0010110}) &= \mathbf{001011} \\ r(\mathbf{0010110}) &= \mathbf{010110} \end{aligned}$$

w		ϵ	**0**	**1**	**00**	**01**	**10**	**11**
$\xi(w)$		0	0	0	0	1	-1	0
w	**000**	**001**	**010**	**011**	**100**	**101**	**110**	**111**
$\xi(w)$	0	1	0	1	-1	0	-1	0

Table 2-0: $\xi(w)$ for $w \in \mathcal{A}_2^{\leq 3}$

$$m(\mathbf{0010110}) = \mathbf{01011}$$
$$R(\mathbf{0010110}) = \mathbf{0110100}$$
$$C(\mathbf{0010110}) = \mathbf{1101001}$$
$$T^7 = \mathbf{0101010}$$
$$CT^7 = \mathbf{1010101}$$

Observation 2.0.0.

$$l \circ r = r \circ l \tag{2.0-0}$$
$$l \circ m = m \circ l \wedge r \circ m = m \circ r \tag{2.0-1}$$
$$R \circ l = r \circ R \wedge R \circ r = l \circ R \tag{2.0-2}$$
$$R \circ m = m \circ R \tag{2.0-3}$$
$$C \circ l = l \circ C \wedge C \circ r = r \circ C \tag{2.0-4}$$
$$C \circ m = m \circ C \tag{2.0-5}$$
$$R \circ C = C \circ R \tag{2.0-6}$$

2.1 The function ξ

– the function ξ

Definition. $\xi\colon \mathcal{A}_2^* \to \{-1, 0, 1\}$ is defined recursively by

$$\xi(w) = \begin{cases} 0 & \text{if } w = \epsilon \\ 1 & \text{if } w \in \{T\}^{\text{even}\geq 0} \\ -1 & \text{if } w \in \{CT\}^{\text{even}\geq 0} \\ \operatorname{sgn}(\xi(l(w)) + \xi(r(w))) & \text{else.} \end{cases}$$

Remark. The function ξ tells on which side of the vertical axis a vertex appears in a de Bruijn graph when drawn as is done in the examples given in this thesis. In figures 2-1 to 2-4 the background is coloured according to the value of ξ evaluated

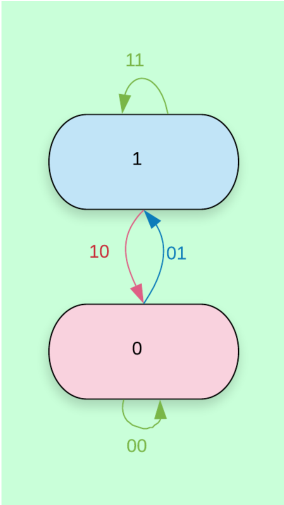

Figure 2-1: The 1-dimensional de Bruijn graph. The vertex **0** and the edge **10** are coloured red indicating that $\phi(\mathbf{0}) = \phi(\mathbf{10}) = -1$ while the vertex **1** and the edge **01** are coloured blue indicating $\phi(\mathbf{1}) = \phi(\mathbf{01}) = 1$. The edges **00**, **11** and the background are coloured green because $\xi(\mathbf{0}) = \xi(\mathbf{1}) = \phi(\mathbf{00}) = \phi(\mathbf{11})$.

---

on the vertices. The vertices w for which $\xi(w) = 0$ are found in the centre. The letter ξ was chosen with the word *x-axis* in mind.

Lemma 2.1.0.

$$\xi \circ C = \xi \circ R = -\xi \tag{2.1-0}$$

Proof. By induction on $\#w$.

$$\xi\big(C(\epsilon)\big) = -\xi(\epsilon) = \xi\big(R(\epsilon)\big) = 0$$

Now assume, the statement holds for $\#w < n$. Let $\#w = n$.

If $w \in \{T\}^{\text{even}} \cup \{CT\}^{\text{even}}$, the statement follows directly from the definition of ξ. Otherwise,

$$\begin{aligned}
(\xi \circ C)(w) &= \operatorname{sgn}\Big((\xi \circ l \circ C)(w) + (\xi \circ r \circ C)(w)\Big) \\
&\overset{(2.0\text{-}4)}{=} \operatorname{sgn}\Big((\xi \circ C \circ l)(w) + (\xi \circ C \circ r)(w)\Big) \\
&\overset{(2.1\text{-}0)}{=} -\operatorname{sgn}\Big((\xi \circ l)(w) + (\xi \circ r)(w)\Big) \\
&= -\xi(w)
\end{aligned}$$

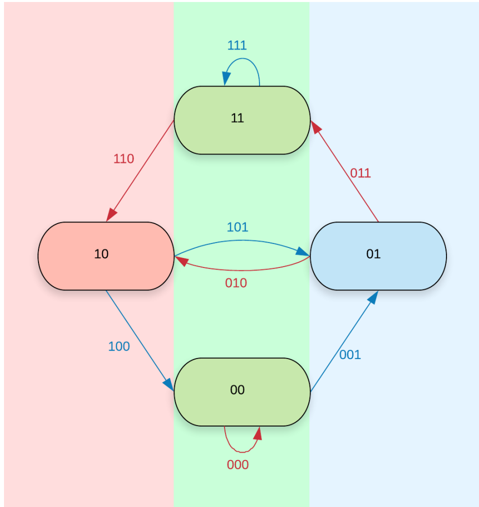

Figure 2-2: The 2-dimensional de Bruijn graph. The vertices and edges are coloured red if $\phi = -1$, green if $\phi = 0$ and blue if $\phi = 1$. The background of the vertex **10** is painted red because $\xi(\mathbf{10}) = -1$. In the centre the background is green because $\xi(\mathbf{00}) = \xi(\mathbf{11}) = 0$ and for the vertex **01** it is blue because $\xi(\mathbf{01}) = 1$.

$$\begin{aligned}
\big(\xi \circ R\big)(w) &= \operatorname{sgn}\Big(\big(\xi \circ l \circ R\big)(w) + \big(\xi \circ r \circ R\big)(w)\Big) \\
&\overset{(2.0\text{-}2)}{=} \operatorname{sgn}\Big(\big(\xi \circ R \circ r\big)(w) + \big(\xi \circ R \circ l\big)(w)\Big) \\
&\overset{(2.1\text{-}0)}{=} -\operatorname{sgn}\Big(\big(\xi \circ r\big)(w) + \big(\xi \circ l\big)(w)\Big) \\
&= -\xi(w).
\end{aligned}$$

Corollary 2.1.1. *Let $w \in \mathcal{A}_2^*$ be a palindrome. Then*

$$\xi(w) = 0. \tag{2.1-1}$$

Proof. Since for a palindrome w, $R(w) = w$, equation (2.1-0) becomes

$$\xi(w) = -\xi(w)$$

which establishes the statement.

Counterexample. **001101** is an example for a string that is not a palindrome although

$$\xi(\mathbf{001101}) = 0.$$

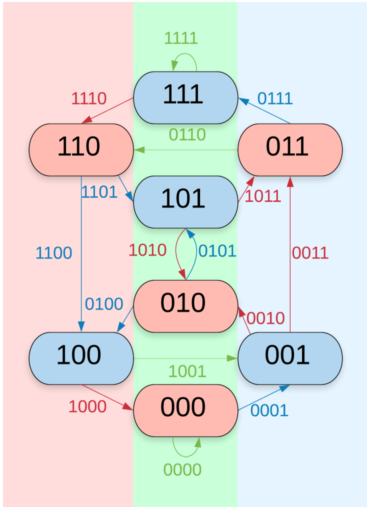

Figure 2-3: The 3-dimensional de Bruijn graph. The vertices and edges are coloured red if $\phi = -1$, green if $\phi = 0$ and blue if $\phi = 1$. The background of the vertices is painted red where $\xi = -1$, green where $\xi = 0$ and blue where $\xi = 1$. Note that since the vertices have odd length, none of them is green.

Corollary 2.1.2. *Let k be odd. Then*

$$\xi\left(T^k\right) = \xi\left(CT^k\right) = 0. \tag{2.1-2}$$

Proof. Since T^k, CT^k are palindromes the statement follows directly from corollary 2.1.1.

Observation 2.1.3. *Let $w \in \mathcal{A}_2^{\geq 1}$.*

$$\xi(w) = 0 \iff w \notin \{T\}^{\text{even}} \cup \left\{CT\right\}^{\text{even}} \wedge \xi\left(l(w)\right) = -\xi\left(r(w)\right) \tag{2.1-3}$$

Corollary 2.1.4. *Let $w \in \mathcal{A}_2^{\geq 1}$ such that $\xi(l(w)), \xi(r(w)) \neq 0$. Then*

$$\xi(w) = 0 \iff \xi\left(l(w)\right) \neq \xi\left(r(w)\right) \tag{2.1-4}$$

Proof. From corollary 2.1.2 it follows that $w \notin \{T\}^{\text{even}} \cup \{CT\}^{\text{even}}$, so the statement follows from observation 2.1.3.

Remark. When $\xi(l(w)), \xi(r(w)) \neq 0$, it is equivalent to say $\xi(l(w)) = -\xi(r(w))$ or $\xi(l(w)) \neq \xi(r(w))$. For aesthetic reasons, the latter formulation will be preferred, also in future occasions.

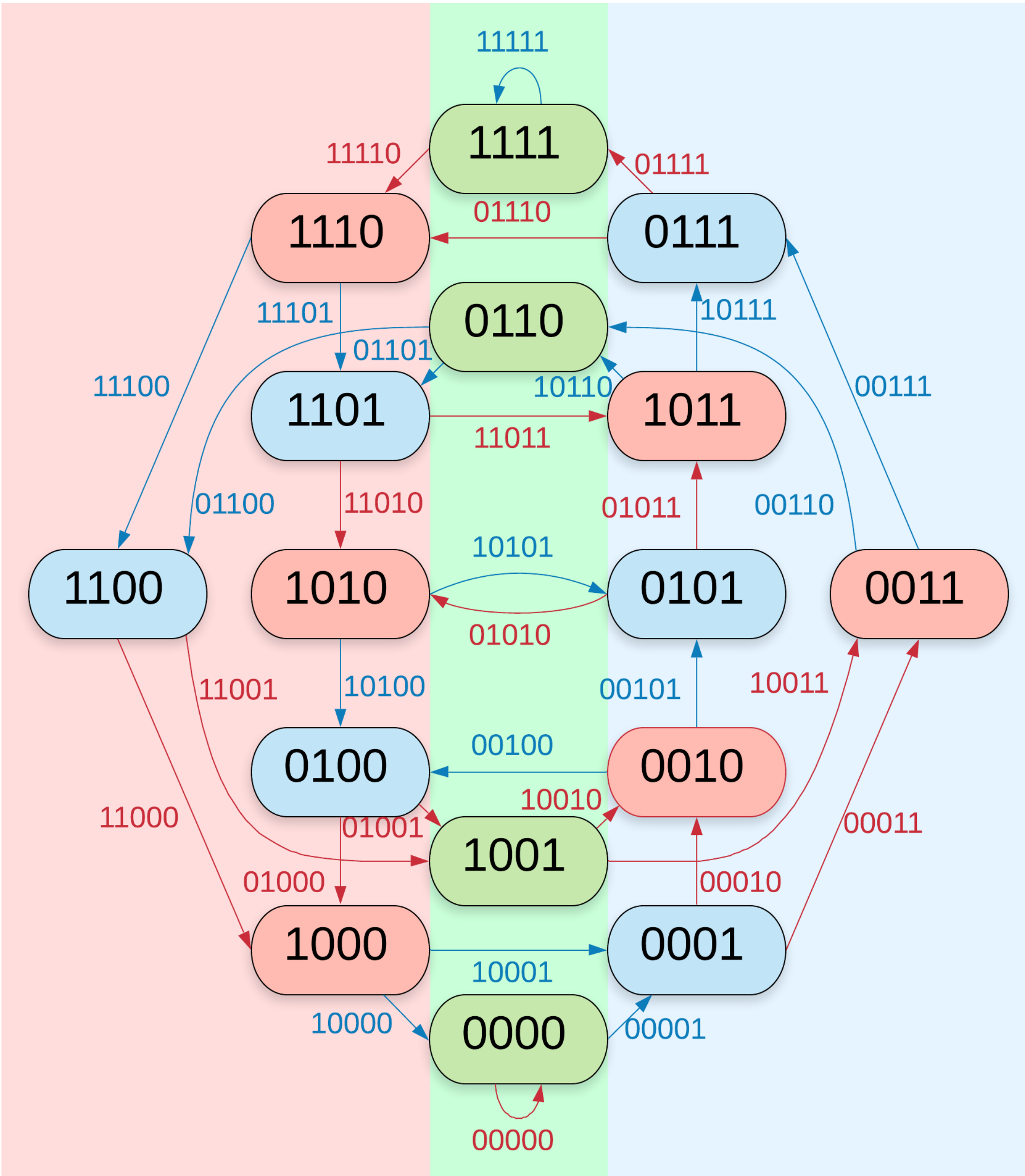

Figure 2-4: The 4-dimensional de Bruijn graph. The vertices and edges are coloured red if $\phi = -1$, green if $\phi = 0$ and blue if $\phi = 1$. The background of the vertices is painted red where $\xi = -1$, green where $\xi = 0$ and blue where $\xi = 1$. Note that since the edges have odd length, none of them is green while the green vertices are exactly those that also have green background.

Lemma 2.1.5. *Let* $w \in \mathcal{A}_2^{\geq 1}$ *such that* $\xi(w) \neq 0$. *Then*

$$\xi\big(l(w)\big) \neq 0 \implies \xi(w) = \xi\big(l(w)\big) \tag{2.1-5}$$
$$\xi\big(r(w)\big) \neq 0 \implies \xi(w) = \xi\big(r(w)\big). \tag{2.1-6}$$

Proof. From $\xi(l(w)) \neq 0 \vee \xi(r(w)) \neq 0$ it follows that $w \overset{(2.1\text{-}2)}{\notin} \{T\}^{\text{even}} \cup \{CT\}^{\text{even}}$, so

$$\xi(w) = \operatorname{sgn}\Big(\xi\big(l(w)\big) + \xi\big(r(w)\big)\Big).$$

Now if one of the summands is non-zero, $\xi(w)$ must either equal that summand or be zero.

Lemma 2.1.6. *Let* $\imath, \jmath \geq 1$. *Then*

$$\xi\big(\mathbf{0}^{\imath}\mathbf{1}^{\jmath}\big) = 1. \tag{2.1-7}$$

Proof. Induction on $\imath + \jmath$. $\xi(\mathbf{01}) = 1$, which establishes the statement for $\imath + \jmath = 2$. Now assume the statement holds for $\imath + \jmath < n \in \mathbb{N}$. Let $\imath + \jmath = n$.

$$\xi\big(\mathbf{0}^{\imath}\mathbf{1}^{\jmath}\big) = \operatorname{sgn}\Big(\xi\Big(\mathbf{0}^{\imath}\mathbf{1}^{\jmath-1}\Big) + \xi\Big(\mathbf{0}^{\imath-1}\mathbf{1}^{\jmath}\Big)\Big),$$

where by the induction hypothesis at least one of the summands is 1 while the other cannot be negative either. The statement follows.

Corollary 2.1.7.

$$\xi\big(\mathbf{1}^{\jmath}\mathbf{0}^{\imath}\big) = -1 \tag{2.1-8}$$

Proof.

$$\xi\big(\mathbf{1}^{\jmath}\mathbf{0}^{\imath}\big) \overset{(2.1\text{-}0)}{=} -\xi\big(\mathbf{0}^{\imath}\mathbf{1}^{\jmath}\big) \overset{(2.1\text{-}7)}{=} -1$$

Lemma 2.1.8. *Let* $k \geq 2$. *Then*

$$\xi\Big(T^k\mathbf{0}\Big) = \begin{cases} 0 & \textit{if } k \textit{ is even} \\ -1 & \textit{if } k \textit{ is odd} \end{cases} \tag{2.1-9}$$

$$\xi\Big(T^k\mathbf{1}\Big) = 1 \tag{2.1-10}$$

$$\xi\Big(CT^k\mathbf{0}\Big) = -1 \tag{2.1-11}$$

$$\xi\Big(CT^k\mathbf{1}\Big) = \begin{cases} 0 & \textit{if } k \textit{ is even} \\ 1 & \textit{if } k \textit{ is odd.} \end{cases} \tag{2.1-12}$$

Proof. By induction on k. Note first, that the statement holds for $k = 2$. Assume then that it holds for a certain $k \geq 2$.

Case k is even:

$$\begin{aligned}
\xi\left(T^{k+1}\mathbf{0}\right) &= \operatorname{sgn}\left(\xi\left(T^{k+1}\right) + \xi\left(CT^{k}\mathbf{0}\right)\right) \overset{(2.1\text{-}2),(2.1\text{-}11)}{=} \operatorname{sgn}\left(0-1\right) = -1 \\
\xi\left(T^{k+1}\mathbf{1}\right) &= \xi\left(T^{k+2}\right) = 1 \\
\xi\left(CT^{k+1}\mathbf{0}\right) &= \xi\left(CT^{k+2}\right) = -1 \\
\xi\left(CT^{k+1}\mathbf{1}\right) &= \operatorname{sgn}\left(\xi\left(CT^{k+1}\right) + \xi\left(T^{k}\mathbf{1}\right)\right) \overset{(2.1\text{-}2),(2.1\text{-}10)}{=} \operatorname{sgn}\left(0+1\right) = 1
\end{aligned}$$

Case k is odd:

$$\begin{aligned}
\xi\left(T^{k+1}\mathbf{0}\right) &= \xi\left(T^{k+2}\right) \overset{(2.1\text{-}2)}{=} 0 \\
\xi\left(T^{k+1}\mathbf{1}\right) &= \operatorname{sgn}\left(\xi\left(T^{k+1}\right) + \xi\left(CT^{k}\mathbf{1}\right)\right) \overset{(2.1\text{-}12)}{=} \operatorname{sgn}\left(1+1\right) = 1 \\
\xi\left(CT^{k+1}\mathbf{0}\right) &= \operatorname{sgn}\left(\xi\left(CT^{k+1}\right) + \xi\left(T^{k}\mathbf{0}\right)\right) \overset{(2.1\text{-}9)}{=} \operatorname{sgn}\left(-1-1\right) = -1 \\
\xi\left(CT^{k+1}\mathbf{1}\right) &= \xi\left(CT^{k+2}\right) \overset{(2.1\text{-}2)}{=} 0
\end{aligned}$$

Lemma 2.1.9. *Let $w \in \mathcal{A}_2^{\geq 1}$ such that $\xi(l(w)) = \xi(r(w)) = 0$. Then*

$$\xi(w) = \ \ 0 \implies w \in \left\{\mathbf{0}\right\}^* \cup \left\{\mathbf{1}\right\}^* \tag{2.1-13}$$
$$\xi(w) = \ \ 1 \implies w \in \{T\}^{\text{even}} \tag{2.1-14}$$
$$\xi(w) = -1 \implies w \in \left\{CT\right\}^{\text{even}}. \tag{2.1-15}$$

Proof. Induction on $\#w$. For $\#w \leq 2$ there are only finitely many cases to consider. Now assume the statements holds for $\#w < n$. Let $\#w = n \geq 3 \wedge \xi(l(w)) = \xi(r(w)) = 0$, which means that

$$\xi\left(l^2(w)\right) = -\xi\left(m(w)\right) = \xi\left(r^2(w)\right).$$

If $m(w) \in \{T\}^{\text{even}} \cup \{CT\}^{\text{even}}$, the statement follows from lemma 2.1.8. Otherwise, the above equation can be rewritten the following way:

$$\begin{aligned}
& \operatorname{sgn}\left(\left(\xi \circ l^3\right)(w) + \left(\xi \circ l^2 \circ r\right)(w)\right) \\
= & -\operatorname{sgn}\left(\left(\xi \circ l^2 \circ r\right)(w) + \left(\xi \circ l \circ r^2\right)(w)\right) \\
= & \ \ \ \operatorname{sgn}\left(\left(\xi \circ l \circ r^2\right)(w) + \left(\xi \circ r^3\right)(w)\right)
\end{aligned}$$

Some playing with the minus sign yields:

$$\begin{aligned} &\operatorname{sgn}\Big(\big(\xi \circ l^3\big)(w) - \big(-\xi \circ l^2 \circ r\big)(w)\Big) \\ =&\operatorname{sgn}\Big(\big(-\xi \circ l^2 \circ r\big)(w) - \big(\xi \circ l \circ r^2\big)(w)\Big) \\ =&\operatorname{sgn}\Big(\big(\xi \circ l \circ r^2\big)(w) - \big(-\xi \circ r^3\big)(w)\Big) \end{aligned}$$

Hence, one of the following three statements must hold:

$$\begin{aligned} &\big(\xi \circ l^3\big)(w) < \big(-\xi \circ l^2 \circ r\big)(w) < \big(\xi \circ l \circ r^2\big)(w) < \big(-\xi \circ r^3\big)(w) \\ \vee\quad &\big(\xi \circ l^3\big)(w) = \big(-\xi \circ l^2 \circ r\big)(w) = \big(\xi \circ l \circ r^2\big)(w) = \big(-\xi \circ r^3\big)(w) \\ \vee\quad &\big(\xi \circ l^3\big)(w) > \big(-\xi \circ l^2 \circ r\big)(w) > \big(\xi \circ l \circ r^2\big)(w) > \big(-\xi \circ r^3\big)(w) \end{aligned}$$

However, since $\operatorname{img}\xi = \{-1, 0, 1\}$, the strict inequalities cannot hold, meaning that the equality does. Then

$$\xi\big(l^2(w)\big) = \xi\big(m(w)\big) = \xi\big(r^2(w)\big) = 0,$$

so by the induction hypothesis

$$l(w), r(w) \in \{\mathbf{0}\}^* \cup \{\mathbf{1}\}^*$$

and since $\#w \geq 3$,

$$w \in \{\mathbf{0}\}^* \cup \{\mathbf{1}\}^*.$$

Remark. The argument used in this proof, considering $l^3, l^2 \circ r, l \circ r^2$ and r^3 of a certain string and using the fact that the functions ξ and ϕ only attain three different values will be important also for the proof of lemma 2.2.9.

Lemma 2.1.10. *Let $w \in \mathcal{A}_2^{\geq 2}$ such that $\xi(w) = 0$. Then*

$$\xi\big(m(w)\big) = 0. \tag{2.1-16}$$

Proof. If $m(w) \in \{T\}^{\text{odd}} \cup \{CT\}^{\text{odd}}$ the statement follows from corollary 2.1.2 so assume it is not. $\xi(w) = 0$ implies that $\xi(l(w)) = -\xi(r(w))$. By lemma 2.1.9 either $w \in \{T\}^{\text{even}} \cup \{CT\}^{\text{even}}$ or $w \in \{\mathbf{0}\}^* \cup \{\mathbf{1}\}^*$ or $\xi(l(w)) = -\xi(r(w)) \neq 0$. However, the first case can be excluded since $\xi(w) = 0$. In the second case, $\xi(m(w)) \overset{(2.1\text{-}1)}{=} 0$. In the third case

$$\operatorname{sgn}\Big(\xi\big(l^2(w)\big) + \xi\big(m(w)\big)\Big) = -\operatorname{sgn}\Big(\xi\big(m(w)\big) + \xi\big(r^2(w)\big)\Big) \neq 0,$$

which can be reformulated to

$$\operatorname{sgn}\Big(\xi\big(m(w)\big) - \Big(-\xi\big(l^2(w)\big)\Big)\Big) = \operatorname{sgn}\Big(\Big(-\xi\big(r^2(w)\big)\Big) - \xi\big(m(w)\big)\Big) \neq 0,$$

so

$$-\xi\big(l^2(w)\big) < \xi\big(m(w)\big) < -\xi\big(r^2(w)\big) \vee -\xi\big(l^2(w)\big) > \xi\big(m(w)\big) > -\xi\big(r^2(w)\big).$$

It follows that $\xi(m(w)) = 0$.

Remark. The number 2.1.11 is omitted in numbering the results to keep the similarity between section 2.1 and 2.2. There is nothing equivalent to say about the function ξ as is said about the function ϕ in corollary 2.2.11.

Lemma 2.1.12. *Let $w \in \mathcal{A}_2^{\geq 2}$ such that $\xi(l^2(w)) = \xi(r^2(w)) \neq 0$. Then*

$$\xi\big(m(w)\big) \neq 0. \tag{2.1-17}$$

Proof. First of all, note that $\#w \geq 4$, as otherwise $\xi(l^2(w)) = 0$. For $\#w = 4$, $w \in \{T^4, CT^4\}$, so $\xi(m(w)) \neq 0$. Assume now that $\#w \geq 5$ and start with considering the case $l^2(w) \in \{T\}^{\text{even}}$. Then $m(w) \in \{CT^{\#w-3}\mathbf{0}, CT^{\#w-3}\mathbf{1}\}$, so $\xi(m(w)) \overset{(2.1\text{-}11),(2.1\text{-}12)}{\neq} 0$. Similarly, the cases $l^2(w) \in \{CT\}^{\text{even}} \vee r^2(w) \in \{T\}^{\text{even}} \cup \{CT\}^{\text{even}}$ can be considered. In all other cases, assume $\xi(m(w)) = 0$. Then

$$\big(\xi \circ l^2 \circ r\big)(w) \overset{(2.1\text{-}3)}{=} \big(-\xi \circ l \circ r^2\big)(w), \tag{2.1-18}$$

so

$$\begin{aligned}
&\xi\big(l^2(w)\big) = \xi\big(r^2(w)\big) \neq 0 \iff \\
&\operatorname{sgn}\Big(\big(\xi \circ l^3\big)(w) + \big(\xi \circ l^2 \circ r\big)(w)\Big) = \operatorname{sgn}\Big(\big(\xi \circ l \circ r^2\big)(w) + \big(\xi \circ r^3\big)(w)\Big) \neq 0 \\
&\overset{(2.1\text{-}18)}{\iff} \\
&\operatorname{sgn}\Big(\big(\xi \circ l^3\big)(w) + \big(\xi \circ l^2 \circ r\big)(w)\Big) = \operatorname{sgn}\Big(\big(\xi \circ r^3\big)(w) - \big(\xi \circ l^2 \circ r\big)(w)\Big) \neq 0 \\
&\implies \big(\xi \circ l^2 \circ r\big)(w) = 0.
\end{aligned}$$

From $(\xi \circ m)(w) = (\xi \circ l^2 \circ r)(w) = (\xi \circ l \circ r^2)(w) = 0$ it follows by lemma 2.1.9 that $m(w) \in \{\mathbf{0}\}^* \cup \{\mathbf{1}\}^*$. However, then

$$\xi\Big(l^2(w)\Big) \leq 0 \leq \xi\Big(r^2(w)\Big),$$

a contradiction.

Remark. One could try to redefine ξ in an attempt to shrink the subset of $\mathcal{A}_2^*$ on which ξ is 0. However lemma 2.1.12 shows that that is not easily possible: Already now ξ is zero only for vertices in the de Bruijn graph that have preceding and succeeding vertices with different ξ-value.

Lemma 2.1.13. *Let* $n \geq 2, w \in \mathcal{A}_2^n$ *such that* $\xi(m(w)) = 0$. *Then*

$$\xi\big(l(w)\big) = \xi\big(r(w)\big) = 0 \iff w \in \big\{\mathbf{0}^n, \mathbf{1}^n\big\} \tag{2.1-19}$$

$$\xi\big(l(w)\big) = \xi\big(r(w)\big) = 1 \iff w \in \Big\{\mathbf{0}T^{n-1}, T^{n-1}\mathbf{1}\Big\} \wedge n \text{ is odd} \tag{2.1-20}$$

$$\xi\big(l(w)\big) = \xi\big(r(w)\big) = -1 \iff w \in \Big\{CT^{n-1}\mathbf{0}, \mathbf{1}CT^{n-1}\Big\} \wedge n \text{ is odd} \tag{2.1-21}$$

$$\xi\big(l(w)\big) = 0 \wedge \xi\big(r(w)\big) = 1 \iff w = \mathbf{0}^{n-1}\mathbf{1} \tag{2.1-22}$$

$$\xi\big(l(w)\big) = 0 \wedge \xi\big(r(w)\big) = -1 \iff w = \mathbf{1}^{n-1}\mathbf{0} \tag{2.1-23}$$

$$\xi\big(l(w)\big) = 1 \wedge \xi\big(r(w)\big) = 0 \iff w = \mathbf{0}\mathbf{1}^{n-1} \tag{2.1-24}$$

$$\xi\big(l(w)\big) = -1 \wedge \xi\big(r(w)\big) = 0 \iff w = \mathbf{1}\mathbf{0}^{n-1}. \tag{2.1-25}$$

Proof.
(2.1-19), (2.1-22), (2.1-23) From $\xi(l(w)) = \xi(m(w)) = 0$ it follows that

$$\xi\Big(l^2(w)\Big) = 0.$$

Hence by lemma 2.1.9 $l(w) \in \{\mathbf{0}^{n-1}, \mathbf{1}^{n-1}\}$

(2.1-24), (2.1-25) Similarly one finds in these cases that $r(w) \in \{\mathbf{0}^{n-1}, \mathbf{1}^{n-1}\}$.

(2.1-20), (2.1-21) Assume $m(w) \notin \{T\}^{\text{odd}} \cup \{CT\}^{\text{odd}}$. Then

$$\xi\Big(l^2(w)\Big) = \xi\Big(r^2(w)\Big) \neq 0$$

which according to lemma 2.1.12 implies that $m(w) \neq 0$, a contradiction. Also, w must not be a palindrome, leaving exactly these four cases.

Remark. Already at this point, $\#\{w \in \mathcal{A}_2^n; \xi(w) = 0\}$ could be investigated. However, that result shall be postponed until section 2.5, as the relevance of that set and its size will be more obvious by then.

Lemma 2.1.14. *Let $w \in \mathcal{A}_2^{\geq 4}$ such that $\xi(l(w)) \neq \xi(m(w)) = -1 \neq \xi(r(w))$. Then $w \in \{T\}^{\text{even}}$.*

Proof. Note that $\xi(l(w)) = \xi(r(w)) = 0$ as anything else would contradict lemma 2.1.5. By lemma 2.1.9

$$w \in \{\mathbf{0}\}^* \cup \{\mathbf{1}\}^* \cup \{T\}^{\text{even}} \cup \{CT\}^{\text{even}},$$

but only $w \in \{T\}^{\text{even}}$ satisfies $\xi(m(w)) = -1$.

Theorem 2.1.15. *Let $k, n \in \mathbb{N}$ and $w \in \mathcal{A}_2^{k+n}$ be such that*

$$w_{[0,n)} = w_{[k,k+n)} \wedge \exists \imath < k \quad \xi\Big(w_{[\imath,\imath+n)}\Big) = 1.$$

Then

$$\exists \jmath < k \quad \xi\Big(w_{[\jmath,\jmath+n)}\Big) = -1.$$

Proof. Induction on n. As it suffices to consider cycles in the n-dimensional de Bruijn graph (instead of any closed walk w represents), there are only finitely many cases to consider for $n \in \{0, 1, 2\}$. Now assume the statement holds for a certain n. Let $w \in \mathcal{A}_2^{k+n+1}$ such that

$$w_{[0,n+1)} = w_{[k,k+n+1)} \wedge \exists \imath < k \quad \xi\Big(w_{[\imath,\imath+n+1)}\Big) = 1.$$

If $w_{[\imath,\imath+n+2)} \in \{T\}^{\text{odd}}$ then $\xi(w_{[\imath+1,\imath+n+2)}) = -1$, which would already establish the statement and if $w_{[\imath,\imath+n+2)} = T^{n+1}\mathbf{1}$ then $\xi(w_{[\imath+2,\imath+n+2)}) = 1$. Otherwise $w_{[\imath,\imath+n+1)} \notin \{T\}^{\text{even}}$ so either $\xi(w_{[\imath,\imath+n)}) = 1 \vee \xi(w_{[\imath+1,\imath+n+1)}) = 1$. In any case either the statement is already established or there is a substring of length n to which ξ assigns the value 1.

Define $\tilde{w}$ by replacing any occurrence of T^{n+2} in w by T^n, meaning that T^{n+2} is not a substring of $\tilde{w}$. Set $\tilde{k} = \#\tilde{w} - (n+1)$. Note that

$$\Big\{\tilde{w}_{[\imath,\imath+n)}; \imath < \tilde{k}\Big\} \subseteq \Big\{w_{[\imath,\imath+n)}; \imath < k\Big\}$$

(in words: any substring of length n of $\tilde{w}$ is also a substring of w), however

$$\Big\{w_{[\imath,\imath+n)}; \imath < k\Big\} \subseteq \Big\{\tilde{w}_{[\imath,\imath+n)}; \imath < \tilde{k}\Big\} \cup \Big\{CT^n\Big\}.$$

(In words: the only substring of length n that has been removed by constructing $\tilde{w}$ from w is CT^n.) Since $\xi(CT^n) \neq 1$, it follows that still $\exists \imath < \tilde{k} \quad \xi(\tilde{w}_{[\imath,\imath+n)}) = 1$. This also shows that $\tilde{k} \geq 1$ and hence $\#\tilde{w} \geq n + 2$.

First, consider the case $\tilde{w}_{[0,n)} \neq \tilde{w}_{[\tilde{k},\tilde{k}+n)}$. That implies $\tilde{w}_{[\tilde{k},\tilde{k}+n+1)} \neq w_{[k,k+n+1)}$, which is possible only if $w_{[k-1,k+n+1)} = T^{n+2}$, so $w_{[0,n+1)} = w_{[k,k+n+1)} = CT^{n+1}$ and hence $\tilde{w}_{[1,n+1)} = \tilde{w}_{[\tilde{k}+1,\tilde{k}+n+1)} = T^n$. By the induction hypothesis,

$$\exists \jmath \quad 1 \leq \jmath < \tilde{k} + 1 \wedge \xi\left(\tilde{w}_{[\jmath,\jmath+n)}\right) = -1. \tag{2.1-26}$$

If $\tilde{w}_{[0,n)} = \tilde{w}_{[\tilde{k},\tilde{k}+n)}$, the induction hypothesis directly gives $\exists \jmath < \tilde{k} \quad \xi(\tilde{w}_{[\jmath,\jmath+n)}) = -1$. If $\jmath = 0$, one can also pick $\jmath = \tilde{k}$, so also in this case equation (2.1-26) holds.

Then by lemma 2.1.14

$$\xi\left(\tilde{w}_{[\jmath-1,\jmath+n)}\right) = -1 \vee \xi\left(\tilde{w}_{[\jmath,\jmath+n+1)}\right) = -1 \vee \tilde{w}_{[\jmath-1,\jmath+n+1)} = T^{n+2},$$

where the third option is excluded by the definition of $\tilde{w}$. The statement now follows from the fact that

$$\left\{\tilde{w}_{[\imath,\imath+n+1)}; \imath < \tilde{k}\right\} \subseteq \left\{w_{[\imath,\imath+n+1)}; \imath < k\right\}.$$

(In words: any substring of length $n+1$ of $\tilde{w}$ is also a substring of w.)

Remark. The statement also holds when switching the rolls of 1 and -1; the proof is analogous.

Using the notion of ξ as defining a left, a centre and a right part of the de Bruijn graph, theorem 2.1.15 tells that there are no non-empty closed walks in the right part. Figure 2-4 does not contain non-trivial closed walks in the blue (or red) highlighted area.

The proof idea is simpler than it seems: A string representing a walk in a certain de Bruijn graph also represents a walk in the de Bruijn graph one dimension lower. If a walk passes a vertex v for which $\xi(v) = -1$ of a graph by the construction of the ξ function it will also do so in a higher dimensional de Bruijn graph – with one notable exception: The vertex CT^n for even n. Take for example the walk **0101** in the 2-dimensional de Bruijn graph. Glancing at figure 2-2 shows that this walk passes both the area with blue and the area with red background. Now consider the same string a walk through the 3-dimensional de Bruijn graph. A look at figure 2-3 tells that the graph does not pass the area with red background. This one exception is what makes the construction of $\tilde{w}$ necessary.

Corollary 2.1.16. *Let $n \in \mathbb{N}$. The sft generated by prohibiting the strings in*

$$\left\{w \in \mathcal{A}_2^n; \xi(w) \neq 1\right\}$$

is empty.

Proof. Assume, there were a point u in the sft. Pick $\imath < \jmath \in \mathbb{N}$ such that $u_{[\imath,\imath+n)} = u_{[\jmath,\jmath+n)}$. By definition of the sft, $\xi(u_{[\imath,\imath+n)}) = 1$. Set $w = u_{[\imath,\jmath+n)}$ and $k = \jmath - \imath$. Then theorem 2.1.15 contradicts the definition of the sft.

w		ϵ	**0**	**1**	**00**	**01**	**10**	**11**
$\phi(w)$		0	-1	1	0	1	-1	0
w	**000**	**001**	**010**	**011**	**100**	**101**	**110**	**111**
$\phi(w)$	-1	1	-1	-1	1	1	-1	1

Table 2-5: $\phi(w)$ for $w \in \mathcal{A}_2^{\leq 3}$

2.2 The function ϕ

Remark. Section 2.2 has the same structure as the previous one. Each result but corollary 2.2.11 can be compared with the result of the same number in section 2.1. Often the proofs are similar.

– *the function* ϕ

Definition. ϕ: $\mathcal{A}_2^* \to \{-1, 0, 1\}$ is defined recursively by

$$\phi(w) = \begin{cases} 0 & \text{if } w = \epsilon \\ -1 & \text{if } w \in \{\mathbf{0}\}^{\text{odd}} \\ 1 & \text{if } w \in \{\mathbf{1}\}^{\text{odd}} \\ \operatorname{sgn}(\phi(r(w)) - \phi(l(w))) & \text{else.} \end{cases}$$

Remark. In figures 2-1 to 2-4 the vertices and edges are coloured according to the value ϕ assigns to them.

Lemma 2.2.0.

$$\phi \circ C = -\phi \qquad \big(\phi \circ R\big)(w) = \big(-1\big)^{\#w+1} \cdot \phi(w) \tag{2.2-0}$$

Proof. By induction on $\#w$.

$$\phi\big(C(\epsilon)\big) = -\phi(\epsilon) = \phi\big(R(\epsilon)\big) = \big(-1\big)^{0+1} \cdot \phi(\epsilon) = 0$$

Now assume, the statement holds for $\#w < n$. Let $\#w = n$.

If $w \in \{\mathbf{0}\}^{\text{odd}} \cup \{\mathbf{1}\}^{\text{odd}}$, the statement follows directly from the definition of ϕ. Otherwise,

$$\begin{aligned} \big(\phi \circ C\big)(w) &= \operatorname{sgn}\Big(\big(\phi \circ r \circ C\big)(w) - \big(\phi \circ l \circ C\big)(w)\Big) \\ &= \operatorname{sgn}\Big(\big(\phi \circ C \circ r\big)(w) - \big(\phi \circ C \circ l\big)(w)\Big) \\ &= -\operatorname{sgn}\Big(\big(\phi \circ r\big)(w) - \big(\phi \circ l\big)(w)\Big) \\ &= -\phi(w) \end{aligned}$$

$$\begin{aligned}(\phi \circ R)(w) &= \operatorname{sgn}\Big((\phi \circ r \circ R)(w) - (\phi \circ l \circ R)(w)\Big)\\ &= \operatorname{sgn}\Big((\phi \circ R \circ l)(w) - (\phi \circ R \circ r)(w)\Big)\\ &= (-1)^n \cdot \operatorname{sgn}\Big((\phi \circ l)(w) - (\phi \circ r)(w)\Big)\\ &= (-1)^{n+1} \cdot \phi(w)\end{aligned}$$

Corollary 2.2.1. *Let* $w \in \mathcal{A}_2^{\text{even}}$ *be a palindrome. Then*

$$\phi(w) = 0.$$

Proof. For a palindrome w of even length, equation (2.2-0) becomes

$$\phi(w) = -\phi(w),$$

which establishes the statement.

Corollary 2.2.2. *Let* k *be even. Then*

$$\phi\left(\mathbf{0}^k\right) = \phi\left(\mathbf{1}^k\right) = 0 \tag{2.2-1}$$

Proof. Since $\mathbf{0}^k, \mathbf{1}^k$ are palindromes, the statement follows directly from corollary 2.2.1.

Observation 2.2.3. *Let* $w \in \mathcal{A}_2^{\geq 1}$.

$$\phi(w) = 0 \iff w \notin \{\mathbf{0}\}^{\text{odd}} \cup \{\mathbf{1}\}^{\text{odd}} \land \phi\big(l(w)\big) = \phi\big(r(w)\big) \tag{2.2-2}$$

Corollary 2.2.4. *Let* $w \in \mathcal{A}_2^{\geq 1}$ *such that* $\phi(l(w)), \phi(r(w)) \neq 0$. *Then*

$$\phi(w) = 0 \iff \phi\big(l(w)\big) = \phi\big(r(w)\big) \tag{2.2-3}$$

Proof. From corollary 2.2.2 it follows that $w \notin \{\mathbf{0}\}^{\text{odd}} \cup \{\mathbf{1}\}^{\text{odd}}$, so the statement follows from observation 2.1.3.

Lemma 2.2.5. *Let* $w \in \mathcal{A}_2^{\geq 1}$ *such that* $\phi(w) \neq 0$. *Then*

$$\phi\big(r(w)\big) \neq 0 \implies \phi(w) = \phi\big(r(w)\big) \tag{2.2-4}$$

$$\phi\big(l(w)\big) \neq 0 \implies \phi(w) \neq \phi\big(l(w)\big). \tag{2.2-5}$$

Proof. From $\phi(r(w)) \neq 0 \vee \phi(l(w)) \neq 0$ it follows that $w \notin \{\mathbf{0}\}^{\text{odd}} \cup \{\mathbf{1}\}^{\text{odd}}$, so

$$\phi(w) = \operatorname{sgn}\Big(\phi\big(r(w)\big) + \Big(-\phi\big(l(w)\big)\Big)\Big).$$

Now if one of the summands is non-zero $\phi(w)$ must either equal that summand or be zero.

Lemma 2.2.6. *Let* $\imath, \jmath \geq 1$. *Then*

$$\phi\big(\mathbf{0}^{\imath}\mathbf{1}^{\jmath}\big) = \big(-1\big)^{\jmath+1}. \tag{2.2-6}$$

Proof. Induction on $\imath + \jmath$. $\phi(\mathbf{01}) = 1$ which establishes the statement for $\imath + \jmath = 2$. Now assume the statement holds for $\imath + \jmath < n \in \mathbb{N}$. Let $\imath + \jmath = n$. Then

$$\phi\big(\mathbf{0}^{\imath}\mathbf{1}^{\jmath}\big) = \operatorname{sgn}\Big(\phi\Big(\mathbf{0}^{\imath-1}\mathbf{1}^{\jmath}\Big) - \phi\Big(\mathbf{0}^{\imath}\mathbf{1}^{\jmath-1}\Big)\Big),$$

where

$$\begin{aligned} \jmath \text{ is even} &\implies \phi\Big(\mathbf{0}^{\imath-1}\mathbf{1}^{\jmath}\Big) \leq 0 \wedge \phi\Big(\mathbf{0}^{\imath}\mathbf{1}^{\jmath-1}\Big) = 1 \\ \jmath \text{ is odd} &\implies \phi\Big(\mathbf{0}^{\imath-1}\mathbf{1}^{\jmath}\Big) = 1 \wedge \phi\Big(\mathbf{0}^{\imath}\mathbf{1}^{\jmath-1}\Big) \leq 0. \end{aligned}$$

In both cases the statement holds.

Corollary 2.2.7.

$$\phi\big(\mathbf{1}^{\jmath}\mathbf{0}^{\imath}\big) = \big(-1\big)^{\imath}$$

Proof.

$$\phi\big(\mathbf{1}^{\jmath}\mathbf{0}^{\imath}\big) \overset{(2.2\text{-}0)}{=} \big(-1\big)^{\imath+\jmath+1} \cdot \phi\big(\mathbf{0}^{\imath}\mathbf{1}^{\jmath}\big) \overset{(2.2\text{-}6)}{=} \big(-1\big)^{\imath+\jmath+1+\jmath+1} = \big(-1\big)^{\imath}$$

Lemma 2.2.8. *Let* $k > 0$. *Then* $\phi(T^k) = (-1)^k$.

Proof. Induction on k. $\phi(T^1) = \phi(\mathbf{0}) = -1$. Now assume the statement holds for a certain k. Then

$$\phi\Big(T^{k+1}\Big) = \operatorname{sgn}\Big(\phi\Big(CT^k\Big) - \phi\Big(T^k\Big)\Big) \overset{(2.2\text{-}0)}{=} \operatorname{sgn}\Big(-2 \cdot \phi\Big(T^k\Big)\Big) = -\phi\Big(T^k\Big) = \big(-1\big)^{k+1}.$$

Lemma 2.2.9.

$$\forall w \in \mathcal{A}_2^{\geq 1} \quad \phi\big(l(w)\big) = \phi\big(r(w)\big) = 0 \implies w \in \big\{\mathbf{0}\big\}^{\text{odd}} \cup \big\{\mathbf{1}\big\}^{\text{odd}}$$

Proof. Induction on $\#w$. For $\#w \leq 2$ there are only finitely many cases. Now assume the statement holds for $\#w < n$. Let $\#w = n \geq 3$ and $\phi(l(w)) = \phi(r(w)) = 0$, which means that

$$\phi\left(l^2(w)\right) = \phi\big(m(w)\big) = \phi\left(r^2(w)\right).$$

By the argument used in the proof of lemma 2.1.9, one gets that in fact

$$\phi\left(l^2(w)\right) = \phi\big(m(w)\big) = \phi\left(r^2(w)\right) = 0,$$

so by the induction hypothesis,

$$l(w), r(w) \in \{\mathbf{0}\}^* \cup \{\mathbf{1}\}^*$$

and since $\#w \geq 3$,

$$w \in \{\mathbf{0}\}^* \cup \{\mathbf{1}\}^*.$$

$\#w$ must be odd because otherwise $\phi(l(w)) \neq 0$.

Lemma 2.2.10. *Let* $w \in \mathcal{A}_2^{\geq 2}$ *such that* $\phi(w) = 0$. *Then*

$$\phi\big(m(w)\big) = 0. \tag{2.2-7}$$

Proof. If $m(w) \in \{\mathbf{0}\}^{\text{even}} \cup \{\mathbf{1}\}^{\text{even}}$ the statement follows from corollary 2.2.2 so assume that it is not. $\phi(w) = 0$ implies that $\phi(l(w)) = \phi(r(w))$. By lemma 2.2.9 either $w \in \{\mathbf{0}\}^{\text{odd}} \cup \{\mathbf{1}\}^{\text{odd}}$ or $\phi(l(w)) = \phi(r(w)) \neq 0$. However, the first case can be excluded since $\phi(w) = 0$. Hence

$$\operatorname{sgn}\left(\phi\big(m(w)\big) - \phi\left(l^2(w)\right)\right) = \operatorname{sgn}\left(\phi\left(r^2(w)\right) - \phi\big(m(w)\big)\right) \neq 0,$$

so

$$\phi\left(l^2(w)\right) < \phi\big(m(w)\big) < \phi\left(r^2(w)\right) \quad \vee \quad \phi\left(l^2(w)\right) > \phi\big(m(w)\big) > \phi\left(r^2(w)\right).$$

In both cases it follows that $\phi(m(w)) = 0$.

Corollary 2.2.11. *Let* $w \in \mathcal{A}_2^{\text{odd}}$. *Then*

$$\phi(w) \neq 0. \tag{2.2-8}$$

Proof. Induction on $\#w$. For $\#w = 1$ there are only finitely many cases to consider. Assume now, the statement is true for $\#w < n$. Let $\#w = n$ and $\phi(w) = 0$. Then $\phi(m(w)) = 0$, so $n - 2$ is odd.

Lemma 2.2.12. *Let $w \in \mathcal{A}_2^{\geq 2}$ such that $\phi(l^2(w)) = \phi(r^2(w)) \neq 0$. Then*

$$\phi\big(m(w)\big) \neq 0.$$

Proof. $\#w = 2 \implies \phi(l^2(w)) = 0$. Now let $\#w \geq 3$ and $\phi(m(w)) = 0$. Then

$$\Big(\phi \circ l^2 \circ r\Big)(w) \overset{(2.2\text{-}2)}{=} \Big(\phi \circ l \circ r^2\Big)(w) \tag{2.2-9}$$

and by corollary 2.2.11 $\#w$ is even so $l^2(w), r^2(w) \notin \{\mathbf{0}\}^{\text{odd}} \cup \{\mathbf{1}\}^{\text{odd}}$. Hence

$$\begin{aligned}
&\phi\Big(l^2(w)\Big) = \phi\Big(r^2(w)\Big) \neq 0 \iff \\
&\operatorname{sgn}\bigg(\Big(\phi \circ l^2 \circ r\Big)(w) - \Big(\phi \circ l^3\Big)(w)\bigg) = \operatorname{sgn}\bigg(\Big(\phi \circ r^3\Big)(w) - \Big(\phi \circ l \circ r^2\Big)(w)\bigg) \neq 0 \\
&\overset{(2.2\text{-}9)}{\iff} \\
&\operatorname{sgn}\bigg(\Big(-\phi \circ l^3\Big)(w) + \Big(\phi \circ l^2 \circ r\Big)(w)\bigg) = \operatorname{sgn}\bigg(\Big(\phi \circ r^3\Big)(w) - \Big(\phi \circ l^2 \circ r\Big)(w)\bigg) \neq 0 \\
&\implies \Big(\phi \circ l^2 \circ r\Big)(w) = 0,
\end{aligned}$$

which is impossible because $\#((l^2 \circ r)(w))$ is odd.

Lemma 2.2.13. *Let $w \in \mathcal{A}_2^2$ such that $\phi(m(w)) = 0$. Then*

$$\phi(w) \neq 0 \iff w \in \Lambda^{\geq 2} = \bigcup_{n \geq 1} \big\{\mathbf{0}^n\mathbf{1}, \mathbf{0}\mathbf{1}^n, \mathbf{1}\mathbf{0}^n, \mathbf{1}^n\mathbf{0}\big\}.$$

Proof. $\Longleftarrow$ is established in lemma 2.2.6 and corollary 2.2.7. To show $\implies$ assume $w \notin \Lambda^{\geq 2}$ but $\phi(m(w)) = 0$. Then $l(w), r(w) \notin \{\mathbf{0}\}^{\text{odd}} \cup \{\mathbf{1}\}^{\text{odd}}$ and by corollary 2.2.11 $\#w$ is even, so neither is w. Hence

$$\begin{aligned}
\phi\big(l(w)\big) &= \operatorname{sgn}\Big(\phi\big(m(w)\big) - \phi\Big(l^2(w)\Big)\Big) = -\phi\Big(l^2(w)\Big) \\
\phi\big(r(w)\big) &= \operatorname{sgn}\Big(\phi\Big(r^2(w)\Big) - \phi\big(m(w)\big)\Big) = \phi\Big(r^2(w)\Big) \\
\phi(w) &= \operatorname{sgn}\Big(\phi\big(r(w)\big) - \phi\big(l(w)\big)\Big) = \operatorname{sgn}\Big(\phi\Big(r^2(w)\Big) + \phi\Big(l^2(w)\Big)\Big).
\end{aligned}$$

Since $\#(l(w))$ and $\#(r(w))$ are odd, both summands are non-zero. However, by lemma 2.2.12 they cannot be equal either. Hence $\phi(w) = 0$.

2.3 Relations between ξ and ϕ

Lemma 2.3.0. *Let* $w \in \mathcal{A}_2^{\geq 2}$ *such that* $\phi(m(w)) = \xi(m(w)) = 0$. *Then*

$$\phi(w) = \xi(w). \tag{2.3-0}$$

Proof. By corollary 2.2.11 $\#w$ is even. Then by lemmata 2.1.13 and 2.2.13

$$\xi(w) \neq 0 \iff \phi(w) \neq 0 \iff w \in \Lambda^{\geq 2}.$$

Set $k = \#w - 1$. According to lemmata 2.1.6, 2.2.6, corollaries 2.1.7 and 2.2.7

$$\xi\left(\mathbf{0}^k\mathbf{1}\right) = \xi\left(\mathbf{0}\mathbf{1}^k\right) = \phi\left(\mathbf{0}^k\mathbf{1}\right) = \phi\left(\mathbf{0}\mathbf{1}^k\right) = 1$$
$$\xi\left(\mathbf{1}^k\mathbf{0}\right) = \xi\left(\mathbf{1}\mathbf{0}^k\right) = \phi\left(\mathbf{1}^k\mathbf{0}\right) = \phi\left(\mathbf{1}\mathbf{0}^k\right) = -1.$$

Corollary 2.3.1. *Let* $w \in \mathcal{A}_2^*$.

$$\phi(w) = 0 \iff \#w \text{ is even} \wedge \xi(w) = 0 \tag{2.3-1}$$

Proof. Induction on $\#w$. For $\#w \leq 1$ there are only finitely many cases to consider. Now assume, the statement holds for $\#w < n$. Let $\#w = n$.

$$\begin{aligned}
\phi(w) = 0 &\overset{(2.2\text{-}8)}{\Longrightarrow} \#w \text{ is even}\\
\phi(w) = 0 &\overset{(2.2\text{-}7)}{\Longrightarrow} \phi\big(m(w)\big) = 0\\
&\overset{(2.3\text{-}1)}{\Longrightarrow} \xi\big(m(w)\big) = 0\\
&\overset{(2.3\text{-}0)}{\Longrightarrow} \xi(w) = \phi(w) = 0\\
\#w \text{ is even} \wedge \xi(w) = 0 &\overset{(2.1\text{-}16)}{\Longrightarrow} \#w \text{ is even} \wedge \xi\big(m(w)\big) = 0\\
&\overset{(2.3\text{-}1)}{\Longrightarrow} \phi\big(m(w)\big) = 0\\
&\overset{(2.3\text{-}0)}{\Longrightarrow} \phi(w) = \xi(w) = 0
\end{aligned}$$

Corollary 2.3.2. *Let* $w \in \mathcal{A}_2^{\geq 1}$ *be such that* $\xi(l(w)), \xi(r(w)) \neq 0$. *Then*

$$\phi\big(l(w)\big) = \phi\big(r(w)\big) \iff \#w \text{ is even} \wedge \xi\big(l(w)\big) \neq \xi\big(r(w)\big) \tag{2.3-2}$$

Proof.

$$\begin{aligned}
\phi\big(l(w)\big) = \phi\big(r(w)\big) &\overset{(2.2\text{-}3)}{\iff} \phi(w) = 0\\
&\overset{(2.3\text{-}1)}{\iff} \#w \text{ is even} \wedge \xi(w) = 0\\
&\overset{(2.1\text{-}4)}{\iff} \#w \text{ is even} \wedge \xi\big(l(w)\big) \neq \xi\big(r(w)\big)
\end{aligned}$$

Remark. By giving a necessary and sufficient condition for the exceptions, corollary 2.3.2 shows that usually $\phi(l(w)) \neq \phi(r(w))$. Later the alternating colouring function ψ shall be introduced as an improvement of ϕ for which the inequality always hold.

Lemma 2.3.3. *Let* $w \in \mathcal{A}_2^{\geq 2}$ *such that*

$$\xi\Big(l^2(w)\Big), \xi\Big(r^2(w)\Big) \neq 0 = \xi\big(m(w)\big). \tag{2.3-3}$$

Then

$$\phi\Big(l^2(w)\Big) = \phi\Big(r^2(w)\Big) \iff \#w \text{ is odd} . \tag{2.3-4}$$

Proof. Note that $l(w), r(w) \overset{(2.3\text{-}3)}{\notin} \{T\}^{\text{even}} \cup \{CT\}^{\text{even}} \cup \{\mathbf{0}\}^* \cup \{\mathbf{1}\}^*$, so

$$\xi\big(l(w)\big) = \operatorname{sgn}\Big(\xi\Big(l^2(w)\Big) + \xi\big(m(w)\big)\Big) \overset{(2.3\text{-}3)}{=} \xi\Big(l^2(w)\Big) \tag{2.3-5}$$

$$\xi\big(r(w)\big) = \operatorname{sgn}\Big(\xi\big(m(w)\big) + \xi\Big(r^2(w)\Big)\Big) \overset{(2.3\text{-}3)}{=} \xi\Big(r^2(w)\Big) \tag{2.3-6}$$

$$\phi\big(l(w)\big) = \operatorname{sgn}\Big(\phi\big(m(w)\big) - \phi\Big(l^2(w)\Big)\Big) \overset{(2.3\text{-}3)}{=} -\phi\Big(l^2(w)\Big) \tag{2.3-7}$$

$$\phi\big(r(w)\big) = \operatorname{sgn}\Big(\phi\Big(r^2(w)\Big) - \phi\big(m(w)\big)\Big) \overset{(2.3\text{-}3)}{=} \phi\Big(r^2(w)\Big). \tag{2.3-8}$$

Hence

$$\begin{aligned}
\phi\Big(l^2(w)\Big) = \phi\Big(r^2(w)\Big) &\overset{(2.3\text{-}7),(2.3\text{-}8)}{\iff} \phi\big(l(w)\big) \neq \phi\big(r(w)\big) \\
&\overset{(2.3\text{-}2)}{\iff} \#w \text{ is odd } \vee \xi\big(l(w)\big) = \xi\big(r(w)\big) \\
&\overset{(2.3\text{-}5),(2.3\text{-}6)}{\iff} \#w \text{ is odd } \vee \xi\Big(l^2(w)\Big) = \xi\Big(r^2(w)\Big) \\
&\overset{(2.1\text{-}17)}{\iff} \#w \text{ is odd.}
\end{aligned}$$

Lemma 2.3.4. *Let* $k, n \in \mathbb{N}, w \in \mathcal{A}_2^{\mathbb{N}}$ *such that* $k \geq 1$ *and*

$$\forall \jmath \leq k \quad \Big(\xi\Big(w_{[\jmath,\jmath+n)}\Big) \neq 0 \iff \jmath \in \big\{0, k\big\}\Big).$$

Then

$$\phi\Big(w_{[0,n)}\Big) = \phi\Big(w_{[k,k+n)}\Big) \iff n \text{ is odd } \wedge \xi\Big(w_{[0,n)}\Big) \neq \xi\Big(w_{[k,k+n)}\Big). \tag{2.3-9}$$

Proof. Since $\xi(w_{[0,n)}) \neq 0$, actually $n \geq 2$. The cases $k \in \{1,2\}$ are covered by corollary 2.3.2 and lemma 2.3.3 respectively. Now let $k \geq 3$. Then by lemma 2.1.9

$$m(w) \in \left\{\mathbf{0}^{k+n-2}, \mathbf{1}^{k+n-2}\right\} \vee n \text{ is odd } \wedge m(w) \in \left\{T^{k+n-2}, CT^{k+n-2}\right\}$$

Taking into account the condition $\xi(w_{[0,n)}), \xi(w_{[k,k+n)}) \neq 0$ gives

$$\begin{aligned} & w \in \left\{\mathbf{10}^{k+n-2}\mathbf{1}, \mathbf{01}^{k+n-2}\mathbf{0}\right\} \\ \vee\ & w \in \left\{\mathbf{0}T^{k+n-2}\mathbf{0}, \mathbf{1}CT^{k+n-2}\mathbf{1}\right\} \wedge k \text{ is even} \wedge n \text{ is odd} \\ \vee\ & w \in \left\{\mathbf{0}T^{k+n-2}\mathbf{1}, \mathbf{1}CT^{k+n-2}\mathbf{0}\right\} \wedge k \text{ is odd } \wedge n \text{ is odd.} \end{aligned}$$

In each of these cases the statement holds.

Remark. The proof idea is to determine those spots in the de Bruijn graph where several edges w for which $\xi(w) = 0$ follow upon each other. By looking at the graphs one gets the impression that happens only around $\mathbf{0}^n$, $\mathbf{1}^n$ and around the centre of odd-dimensional de Bruijn graphs, which lemma 2.1.9 confirms. In all other cases the statement is already established by corollary 2.3.2 and lemma 2.3.3.

2.4 The alternating colouring function ψ

Definition. Define $\psi : \mathcal{A}_2^* \to \{-1, 0, 1\}$ by

$$\psi(w) = \left(\xi(w)\right)^{\#w} \cdot \phi(w). \tag{2.4-0}$$

ψ is called the *alternating colouring function*. – *alternating colouring function* ψ

$w \in \mathcal{A}_2^*$ is called *colourable* if $\psi(w) \neq 0$. For $n \geq 2$ the sft generated by prohibiting the strings in $\{w \in \mathcal{A}_2^n; \psi(w) = 0\}$ is called the *maximal n-colourable shift*. Any subshift of the maximal n-colourable shift is called *n-colourable*. – *colourable string* – *maximal n-colourable shift* – *n-colourable shift*

Remark. In figures 2-7 to 2-10 the vertices and edges are coloured according to the value ψ assigns to them. Colourable are those vertices and edges that are coloured blue or red.

The maximal n-colourable shift is the edge shift of the $(n-1)$-dimensional de Bruijn graph of which all the non-colourable edges (the green ones in figures 2-1 to 2-4 that is) have been removed. Equivalently it is the vertex shift of the n-dimensional de Bruijn graph of which all non-colourable vertices have been removed. Note however that such a graph does not fit the definition of a de Bruijn subgraph because it is not spanning.

w		ϵ	**0**	**1**	**00**	**01**	**10**	**11**
$\psi(w)$		0	0	0	0	1	-1	0
w	**000**	**001**	**010**	**011**	**100**	**101**	**110**	**111**
$\psi(w)$	0	1	0	-1	-1	0	1	0

Table 2-6: The values $\psi(w)$ of the alternating colouring for $w \in \mathcal{A}_2^{\leq 3}$

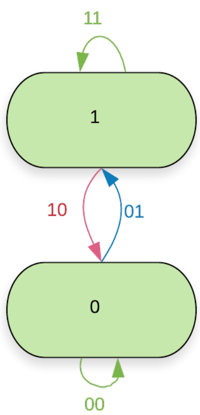

Figure 2-7: The 1-dimensional de Bruijn graph. The edges **01** and **10** are coloured blue and red respectively indicating that $\psi(\mathbf{01}) = 1 \wedge \psi(\mathbf{10}) = -1$. The other two edges and the vertices are coloured green because ψ is 0 there.

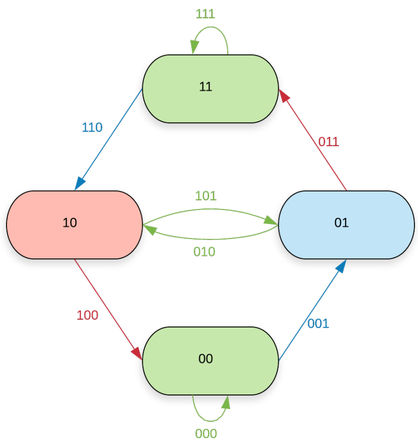

Figure 2-8: The 2-dimensional de Bruijn graph. The vertices and edges are coloured red if $\psi = -1$, green if $\psi = 0$ and blue if $\psi = 1$.

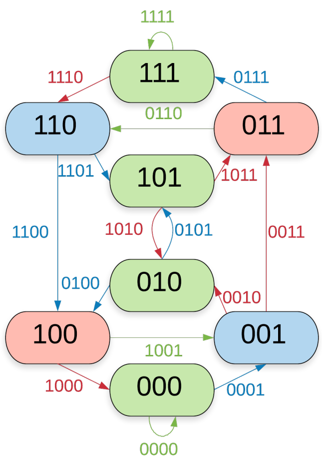

Figure 2-9: The 3-dimensional de Bruijn graph. The vertices and edges are coloured red if $\psi = -1$, green if $\psi = 0$ and blue if $\psi = 1$.

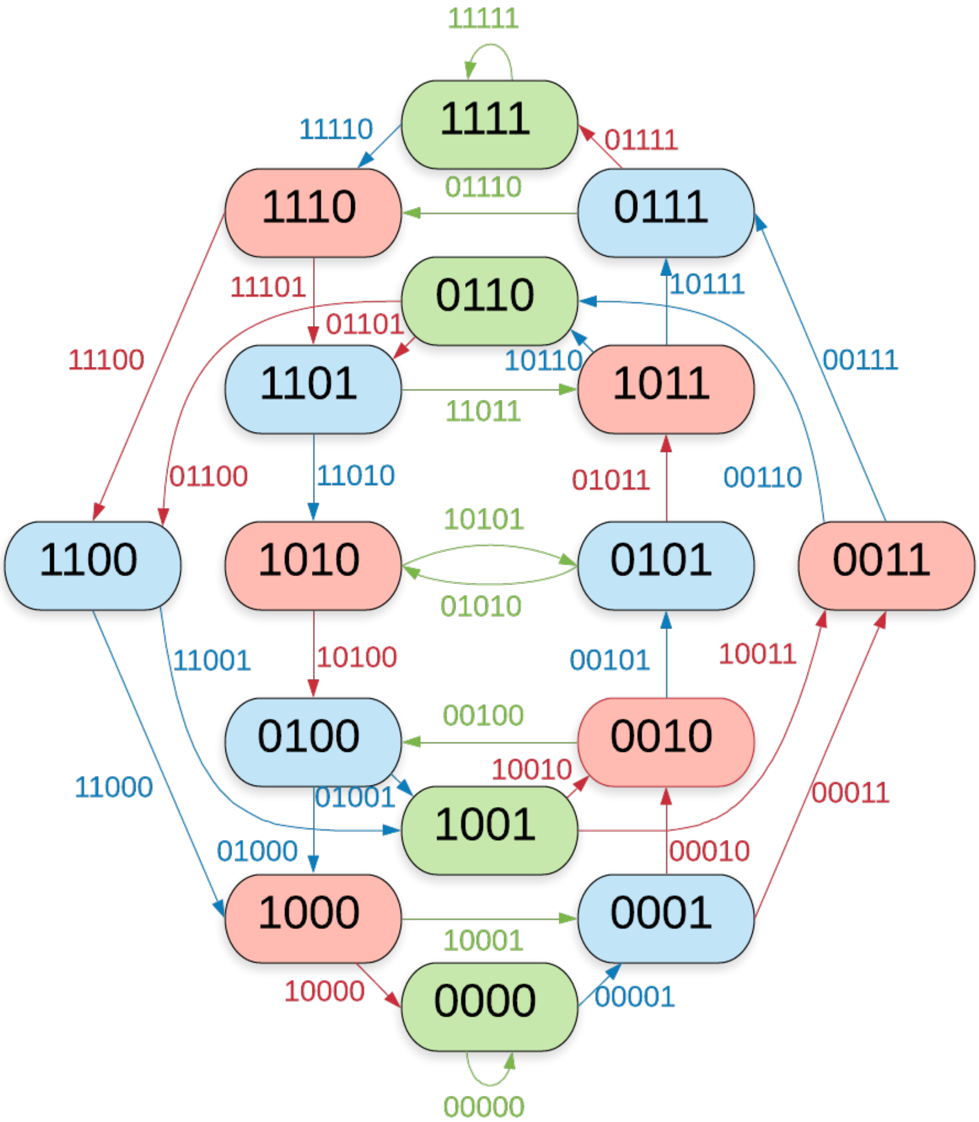

Figure 2-10: The 4-dimensional de Bruijn graph. The vertices and edges are coloured red if $\psi = -1$, green if $\psi = 0$ and blue if $\psi = 1$.

Lemma 2.4.0.

$$\big(\psi \circ C\big)(w) = \big(-1\big)^{\#w+1} \cdot \psi(w) \qquad\qquad \psi \circ R = -\psi \tag{2.4-1}$$

Proof.

$$\begin{array}{rl|rl} & \big(\psi \circ C\big)(w) & & \big(\psi \circ R\big)(w) \\ \overset{(2.4\text{-}0)}{=} & \Big(\xi\big(C(w)\big)\Big)^{\#w} \cdot \phi\big(C(w)\big) & \overset{(2.4\text{-}0)}{=} & \Big(\xi\big(R(w)\big)\Big)^{\#w} \cdot \phi\big(R(w)\big) \\ \overset{(2.1\text{-}0),(2.2\text{-}0)}{=} & \big(-\xi(w)\big)^{\#w} \cdot -\phi(w) & \overset{(2.1\text{-}0),(2.2\text{-}0)}{=} & \big(-\xi(w)\big)^{\#w} \cdot \big(-1\big)^{\#w+1} \cdot \phi(w) \\ \overset{(2.4\text{-}0)}{=} & \big(-1\big)^{\#w+1} \cdot \psi(w) & \overset{(2.4\text{-}0)}{=} & -\psi(w) \end{array}$$

Lemma 2.4.1. *Let* $w \in \mathcal{A}_2^*$*. Then*

$$\psi(w) = 0 \iff \xi(w) = 0. \tag{2.4-2}$$

Proof.

$$\psi(w) = 0 \overset{(2.4\text{-}0)}{\iff} \xi(w) = 0 \vee \phi(w) = 0 \overset{(2.3\text{-}1)}{\iff} \xi(w) = 0$$

Remark. Lemma 2.4.1 will often be used without explicit reference. Saying that w is colourable should always be understood as $\xi(w), \psi(w) \neq 0$. To show colourability, of course the easier condition $\xi(w) \neq 0$ will be used.

Corollary 2.4.2. *Let* $w \in \mathcal{A}_2^{\text{even}}$*. Then*

$$\psi(w) = \phi(w). \tag{2.4-3}$$

Proof. The statement follows directly from lemma 2.4.1 and the definition of ψ.

Theorem 2.4.3. *Let* $k, n \in \mathbb{N}, w \in \mathcal{A}_2^{\mathbb{N}}$ *such that* $k \geq 1$ *and*

$$w_{[J,J+n)} \text{ is colourable } \iff J \in \big\{0, k\big\}.$$

Then

$$\psi\Big(w_{[0,n)}\Big) \neq \psi\Big(w_{[k,k+n)}\Big). \tag{2.4-4}$$

0100111110001101...

Figure 2-11: The beginning of an infinite string $w \in \mathcal{A}_2^{\mathbb{N}}$. For $k \in \mathbb{N}$ the character on place k has been coloured red if $\psi(w_{[k,k+4)}) = -1$, green if $\psi(w_{[k,k+4)}) = 0$ and blue if $\psi(w_{[k,k+4)}) = 1$. The first character is blue because $\psi(\mathbf{0100}) = 1$; the second one is green because $\psi(\mathbf{1001}) = 0$. The last three characters are printed black because their colour depends on the upcoming characters that are not given here. Ignoring the green characters, always a blue character follows on a red and vice versa. The fact that there are green characters means that this w cannot be a point in a 4-colourable shift.

Proof. The statement can be seen as a corollary to lemma 2.3.4. First let n be even. Then

$$\psi\left(w_{[0,n)}\right) \overset{(2.4\text{-}3)}{=} \phi\left(w_{[0,n)}\right) \overset{(2.3\text{-}9)}{\neq} \phi\left(w_{[k,k+n)}\right) \overset{(2.4\text{-}3)}{=} \psi\left(w_{[k,k+n)}\right).$$

Now let n be odd. Then

$$\begin{aligned}\psi\left(w_{[0,n)}\right) &\overset{(2.4\text{-}0)}{=} \xi\left(w_{[0,n)}\right) \cdot \phi\left(w_{[0,n)}\right)\\ &\overset{(2.3\text{-}9)}{\neq} \xi\left(w_{[k,k+n)}\right) \cdot \phi\left(w_{[k,k+n)}\right)\\ &\overset{(2.4\text{-}0)}{=} \psi\left(w_{[k,k+n)}\right).\end{aligned}$$

Remark. Theorem 2.4.3 states that in an infinite string, a colourable substring always has a different ψ-colour than the preceding colourable substring of the same length – no matter how many non-colourable substrings there are in between.

Corollary 2.4.4. *Let $w \in \mathcal{A}_2^{\geq 1}$ be such that $l(w), r(w)$ are colourable. Then $\psi(l(w)) \neq \psi(r(w))$.*

Proof. The statement follows from theorem 2.4.3 by regarding w as the beginning of an infinite string and setting $k = 1, n = \#w - 1$.

Remark. Compare corollary 2.4.4 with corollary 2.3.2. While there were some special cases in which $\phi(l(w)) \neq \phi(r(w))$ would not hold, $\psi(l(w)) \neq \psi(r(w))$ always does.

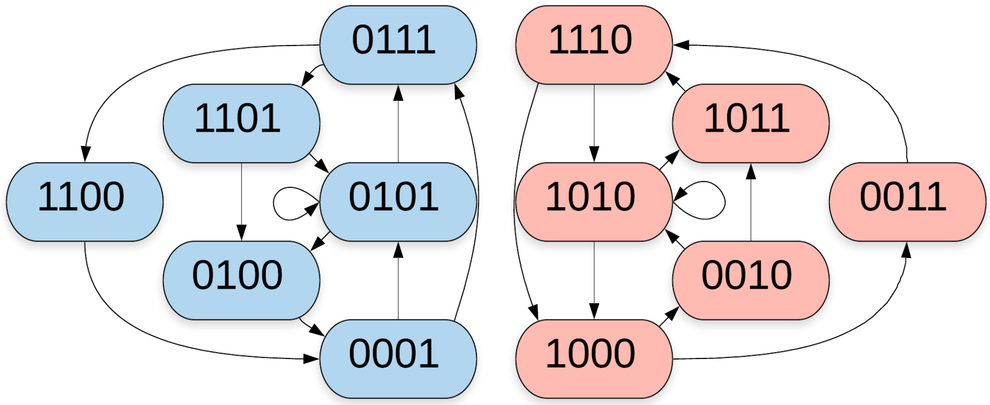

Figure 2-12: The 2nd power graph of the 4-dimensional de Bruijn graph. The vertices whose ξ-value is zero are omitted as are the edges that represent walks which would have passed an omitted edge. The vertices are coloured red if $\xi = -1$ and blue if $\xi = 1$.

Remark. Consider a de Bruijn graph of which all non-colourable vertices have been removed. Corollary 2.4.4 states that the colour of the vertices a walk passes alternates between blue and red. For the resulting vertex shift that result will be expressed in corollary 2.4.7.

In the even powers of such a graph there are only walks from vertices of one colour to vertices of the same colour. Figure 2-12 shows the 2nd power of the 4-dimensional de Bruijn subgraph without non-colourable edges.

Corollary 2.4.5. *Let* $k, n \in \mathbb{N}, w \in \mathcal{A}_2^{\mathbb{N}}$ *such that* $w_{[0,n)}, w_{[k,k+n)}$ *are colourable. Set*

$$z = \#\Big\{ \jmath < k; w_{[\jmath,\jmath+n)} \textit{ is not colourable} \Big\}.$$

Then

$$\psi\Big(w_{[k,k+n)}\Big) = (-1)^{k-z} \cdot \psi\Big(w_{[0,n)}\Big). \tag{2.4-5}$$

Proof. Strong induction on k. For $k = 0$ the statement is trivial. Now assume the statement holds whenever $k < K$ for a certain $K \in \mathbb{N}$. Set

$$\begin{aligned} Z = &\quad \#\Big\{ \jmath < K; w_{[\jmath,\jmath+n)} \text{ is not colourable} \Big\}, \\ k = &\max\Big\{ \jmath < K; w_{[\jmath,\jmath+n)} \text{ is colourable} \Big\}, \end{aligned}$$

000101110001110 1...

Figure 2-13: The beginning of a point w in a 4-colourable shift. For $k \in \mathbb{N}$ the character on place k has been coloured blue if $\psi(w_{[k,k+4)}) = 1$ and red if $\psi(w_{[k,k+4)}) = -1$. The sequence alternates between blue and red, a visualisation of the alternating colouring.

$$z = \quad \#\Big\{ \jmath < k; w_{[\jmath,\jmath+n)} \text{ is not colourable} \Big\}.$$

Note that $Z - z = K - k - 1$, so

$$k - z = K - Z - 1. \tag{2.4-6}$$

Hence

$$\begin{aligned}
\psi\Big(w_{[K,K+n)}\Big) &\overset{(2.4\text{-}4)}{=} - \psi\Big(w_{[k,k+n)}\Big) \\
&\overset{(2.4\text{-}4)}{=} - (-1)^{k-z} \cdot \psi\Big(w_{[0,n)}\Big) \\
&\overset{(2.4\text{-}6)}{=} (-1)^{K-Z} \cdot \psi\Big(w_{[0,n)}\Big).
\end{aligned}$$

Corollary 2.4.6. *Take k, n, w, z as in corollary 2.4.5.*
Let $\psi(w_{[0,n)}) = \psi(w_{[k,k+n)})$. Then

$$k \text{ is even} \iff z \text{ is even.}$$

Proof. By corollary 2.4.5 $(-1)^{k-z} = 1$.

Corollary 2.4.7. *Let $n \in \mathbb{N}$ and w be a point in an n-colourable shift. Then*

$$\forall k \in \mathbb{N} \quad \psi\Big(w_{[k,k+n)}\Big) = (-1)^k \cdot \psi\Big(w_{[0,n)}\Big). \tag{2.4-7}$$

Proof. The statement follows directly from corollary 2.4.5 by noting that for an n-colourable shift, $z = 0$.

Remark. Corollary 2.4.7 motivates the term *alternating colouring*.

Corollary 2.4.8. *Let $k, n \in \mathbb{N}$ and w be a point in an n-colourable shift such that $w_{[0,n)} = w_{[k,k+n)}$. Then k is even.*

Proof. By corollary 2.4.7 $(-1)^k = 1$.

Remark. Corollary 2.4.8 states that any closed walk among the colourable edges of a de Bruijn graph has even length.

Corollary 2.4.9. *Let $n \in \mathbb{N}$. A non-empty, n-colourable shift is not mixing.*

Proof. For any natural number there is a larger odd number, but by corollary 2.4.8 there is no walk of odd length from a word back to itself.

2.5 Counting the non-colourable strings

Observation 2.5.0. *A string of length ≤ 5 is colourable if and only if it is not a palindrome.*

Lemma 2.5.1. *Let $n \geq 4$. Then*

$$\#\Big\{w \in \mathcal{A}_2^n; \xi\big(m(w)\big) = 0 \wedge \xi(w) \neq 0\Big\} = 6 - 2 \cdot \big(-1\big)^n \tag{2.5-0}$$

Proof. The statement can be seen as a corollary to lemma 2.1.13. Set

$$W = \Big\{w \in \mathcal{A}_2^n; \xi\big(m(w)\big) = 0\Big\}.$$

Then

$$\begin{aligned}
& \#\Big\{w \in \mathcal{A}_2^n; \xi\big(m(w)\big) = 0 \wedge \xi(w) \neq 0\Big\} \\
= \quad & \#\big\{w \in W; \xi(w) \neq 0\big\} \\
= \quad & \#\big\{w \in W; \xi\big(l(w)\big) = 0 \wedge \xi\big(r(w)\big) \neq 0\big\} \\
& + \#\big\{w \in W; \xi\big(l(w)\big) \neq 0 \wedge \xi\big(r(w)\big) = 0\big\} \\
& + \#\big\{w \in W; \xi\big(l(w)\big) = \xi\big(r(w)\big) \neq 0\big\} \\
\overset{(2.1\text{-}22)}{=} & 2 \overset{(2.1\text{-}24)}{+} 2 \overset{(2.1\text{-}20),(2.1\text{-}21)}{+} 2 \cdot \Big(1 - \big(-1\big)^n\Big) \\
= \quad & 6 - 2 \cdot \big(-1\big)^n.
\end{aligned}$$

Corollary 2.5.2. *For $n \in \mathbb{N}$ set $K_n = \#\{w \in \mathcal{A}_2^n; \xi(w) = 0\}$. Let $n \geq 4$. Then*

$$K_n = 4 \cdot K_{n-2} - 6 + 2 \cdot \big(-1\big)^n. \tag{2.5-1}$$

n	0	1	2	3	4	5	6	7	8	9	10	11
J_n	0	1	1	3	5	11	21	43	85	171	341	683

Table 2-14: The first Jacobsthal numbers

Proof.

$$\begin{aligned} K_n &\overset{(2.1\text{-}16)}{=} \#\Big\{w \in \mathcal{A}_2^n; \xi\big(m(w)\big) = 0\Big\} \\ &\quad - \#\Big\{w \in \mathcal{A}_2^n; \xi\big(m(w)\big) = 0 \land \xi(w) \neq 0\Big\} \\ &\overset{(2.5\text{-}0)}{=} 4 \cdot K_{n-2} - 6 + 2 \cdot \big(-1\big)^n. \end{aligned}$$

Definition. For $n \in \mathbb{N}$ the *Jacobsthal number* J_n is defined as — *Jacobsthal number*

$$J_n = \frac{2^n - (-1)^n}{3}. \tag{2.5-2}$$

Fact 2.5.3. *Let* $n \in \mathbb{N}$.

$$J_{n+1} = 2 \cdot J_n + \big(-1\big)^n \tag{2.5-3}$$

$$J_{n+2} = J_{n+1} + 2 \cdot J_n \tag{2.5-4}$$

Remark. The definition of the Jacobsthal numbers, the values in table 2-14 and fact 2.5.3 have been taken from Wikipedia contributors 2020c.

Theorem 2.5.4. *Let* $n \in \mathbb{N}$.

$$K_{n+2} = 2 \cdot \big(J_n + 1\big). \tag{2.5-5}$$

Proof. By induction on n. For $n \leq 1$ there are only two cases to consider. Now assume the statement holds for a certain n.

$$\begin{aligned} K_{n+4} &\overset{(2.5\text{-}1)}{=} 4 \cdot K_{n+2} - 6 + 2 \cdot \big(-1\big)^n \\ &\overset{(2.5\text{-}5)}{=} 8 \cdot \big(J_n + 1\big) - 6 + 2 \cdot \big(-1\big)^n \\ &= 2 \cdot \Big(2 \cdot J_n + \big(-1\big)^n\Big) + 4 \cdot J_n + 2 \\ &\overset{(2.5\text{-}3)}{=} 2 \cdot J_{n+1} + 4 \cdot J_n + 2 \\ &\overset{(2.5\text{-}4)}{=} 2 \cdot \big(J_{n+2} + 1\big) \end{aligned}$$

n	0	1	2	3	4	5	6	7	8	9	10	11
K_n	1	2	2	4	4	8	12	24	44	88	172	344

Table 2-15: The number K_n of non-colourable strings in $\mathcal{A}_2^n$

Corollary 2.5.5. *Let $n \geq 2$. Then*

$$n \text{ is even} \implies K_n = \frac{2^n + 8}{6} \tag{2.5-6}$$

$$n \text{ is odd} \implies K_n = \frac{2^n + 16}{6}. \tag{2.5-7}$$

Proof.

$$K_n \overset{(2.5\text{-}5)}{=} 2 \cdot \left(J_{n-2} + 1\right) \overset{(2.5\text{-}2)}{=} 2 \cdot \left(\frac{2^{n-2} - (-1)^n}{3} + 1\right) = \frac{2^n + 12 - 4 \cdot (-1)^n}{6}$$

Corollary 2.5.6. *Let n be even. Then*

$$K_{n+1} = 2 \cdot K_n. \tag{2.5-8}$$

Proof.

$$K_{n+1} \overset{(2.5\text{-}7)}{=} \frac{2^{n+1} + 16}{6} = 2 \cdot \frac{2^n + 8}{6} \overset{(2.5\text{-}6)}{=} 2 \cdot K_n$$

Corollary 2.5.7.

$$\lim_{n \to \infty} \frac{K_n}{2^n} = \frac{1}{6}$$

Proof. This follows immediately from corollary 2.5.5.

Remark. Since there are 2^n strings of length n, corollary 2.5.7 tells that in the limit 1 out of 6 strings is not colourable.

Corollary 2.5.8. *Let $n \geq 2$. Then*

$$K_{n+4} = 5 \cdot K_{n+2} - 4 \cdot K_n.$$

Proof. Let $c \in \mathbb{N}$. The statement follows from corollary 2.5.5 and the observation that

$$5 \cdot \frac{2^{n+2} + c}{6} - 4 \cdot \frac{2^n + c}{6} = \frac{5 \cdot 2^{n+2} - 2^{n+2} + c}{6} = \frac{2^{n+4} + c}{6}.$$

2.6 Colourability sources and sinks

Definition. Let $w \in \mathcal{A}_2^*$. If $w\mathbf{0}, w\mathbf{1}$ are not colourable, w is called a *colourability sink*; if $\mathbf{0}w, \mathbf{1}w$ are not colourable, w is called a *colourability source*. ϵ is called the *trivial* colourability sink and source.

– colourability sink
– colourability source
– trivial colourability sink and source

Observation 2.6.0. *The reverse of a colourability source is a colourability sink and vice versa.*

Example. **010001** is a colourability sink. According to observation 2.6.0 **100010** is a colourability source.

Corollary 2.6.1. *Let $n \in \mathbb{N}$. There are equally many colourability sinks as sources in $\mathcal{A}_2^n$.*

Proof. If $w \in \mathcal{A}_2^n$ is a colourability sink, $R(w)$ is a colourability source and vice versa. Since R is bijective, the statement follows.

Lemma 2.6.2. *If $w \in \mathcal{A}_2^*$ is a colourability sink $\mathbf{0}w, \mathbf{1}w$ are colourable. If w is a colourability source $w\mathbf{0}, w\mathbf{1}$ are colourable.*

Proof. Assume the contrary, say $\xi(w\mathbf{0}) = \xi(\mathbf{0}w) = 0$. Then by lemma 2.1.9 $\mathbf{0}w\mathbf{0} \in \{\mathbf{0}\}^*$, so $w \in \{\mathbf{0}\}^*$. Hence $\xi(\mathbf{1}w), \xi(w\mathbf{1}) \overset{(2.1\text{-}7),(2.1\text{-}8)}{\neq} 0$, a contradiction. Similarly if $\xi(w\mathbf{1}) = \xi(\mathbf{1}w) = 0$ one gets that $\xi(\mathbf{0}w), \xi(w\mathbf{0}) \neq 0$.

Corollary 2.6.3. *The only string that is both a colourability sink and a colourability source is ϵ.*

Proof. Any non-trivial colourability sink is by lemma 2.6.2 not a colourability source.

Remark. Corollary 2.6.3 means that removing the non-colourable edges of a de Bruijn graph does not leave isolated vertices.

Lemma 2.6.4. *Any non-trivial colourability sink is colourable.*

Proof. Let $w \in \mathcal{A}_2^*$ not be colourable.

$$\xi(w) = \xi\big(w\mathbf{0}\big) = 0 \overset{(2.1\text{-}3)}{\Longrightarrow} \xi(w) = \xi\big(r(w)\mathbf{0}\big) = \xi\big(w\mathbf{0}\big) = 0$$
$$\overset{(2.1\text{-}13)}{\Longrightarrow} w\mathbf{0} \in \big\{\mathbf{0}\big\}^* \cup \big\{\mathbf{1}\big\}^*$$

$$\implies w \in \{\mathbf{0}\}^*$$

but similarly

$$\xi(w) = \xi(w\mathbf{1}) = 0 \implies w \in \{\mathbf{1}\}^*,$$

so $w = \epsilon$.

Corollary 2.6.5. *Any non-trivial colourability source is colourable.*

Proof. Since the reverse of a colourability source is a colourability sink, the statement follows directly from lemmata 2.1.0 and 2.6.4.

Lemma 2.6.6. *If $w \in \mathcal{A}_2*$ is a colourability sink,*

$$\xi(\mathbf{0}w) = \xi(\mathbf{1}w) = \xi(w) \tag{2.6-0}$$
$$\phi(\mathbf{0}w) = \phi(\mathbf{1}w) = \phi(w) \tag{2.6-1}$$
$$\psi(\mathbf{0}w) = \psi(\mathbf{1}w) = \xi(w) \cdot \psi(w). \tag{2.6-2}$$

If w is a colourability source,

$$\xi(w\mathbf{0}) = \xi(w\mathbf{1}) = \xi(w) \tag{2.6-3}$$
$$\phi(w\mathbf{0}) = \phi(w\mathbf{1}) = -\phi(w) \tag{2.6-4}$$
$$\psi(w\mathbf{0}) = \psi(w\mathbf{1}) = -\xi(w) \cdot \psi(w). \tag{2.6-5}$$

Proof. Consider lemma 2.6.2, lemma 2.6.4 and corollary 2.6.5. For (2.6-0) and (2.6-3) combine them with lemma 2.1.5, for (2.6-1) and (2.6-4) with lemma 2.2-5. The following establishes (2.6-2):

$$\left.\begin{array}{l}\psi(\mathbf{0}w) = \left(\xi(\mathbf{0}w)\right)^{\#w+1} \cdot \phi(\mathbf{0}w) \overset{(2.6\text{-}0)}{=} \\ \psi(\mathbf{1}w) = \left(\xi(\mathbf{1}w)\right)^{\#w+1} \cdot \phi(\mathbf{1}w) \overset{(2.6\text{-}1)}{=}\end{array}\right. \left(\xi(w)\right)^{\#w+1} \cdot \phi(w) = \xi(w) \cdot \psi(w)$$

and the following (2.6-5):

$$\left.\begin{array}{l}\psi(w\mathbf{0}) = \left(\xi(w\mathbf{0})\right)^{\#w+1} \cdot \phi(w\mathbf{0}) \overset{(2.6\text{-}3)}{=} \\ \psi(w\mathbf{1}) = \left(\xi(w\mathbf{1})\right)^{\#w+1} \cdot \phi(w\mathbf{1}) \overset{(2.6\text{-}4)}{=}\end{array}\right. \left(\xi(w)\right)^{\#w+1} \cdot \left(-\phi(w)\right) = -\xi(w) \cdot \psi(w).$$

Theorem 2.6.7. *Let $w \in \mathcal{A}_2^*$. Then*

$$w\mathbf{0}, w\mathbf{1} \textit{ are colourable} \implies \psi(w\mathbf{0}) = \psi(w\mathbf{1})$$
$$\mathbf{0}w, \mathbf{1}w \textit{ are colourable} \implies \psi(\mathbf{0}w) = \psi(\mathbf{1}w)$$

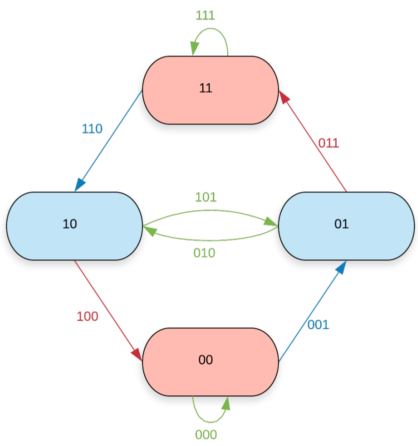

Figure 2-16: The 2-dimensional de Bruijn graph. As in figure 2-8 the edges are coloured according to the value ψ assigns to them. The colour of the vertices coincides with the colour of the incoming edges and differs from the colour of the outgoing ones.

Proof. If w is a colourability sink or source, the statement is proven in lemma 2.6.6. Otherwise it follows from corollary 2.4.4.

Remark. Theorem 2.6.7 makes it possible to define a function that assigns to the vertices of a de Bruijn graph a colour that corresponds to the colour of the incoming edges while differing from the colour of the outgoing ones. Lemma 2.2.5 and corollary 2.2.11 show that for odd-dimensional de Bruijn graphs that function is just ϕ. Figures 2-16 and 2-17 give examples for even-dimensional de Bruijn graphs.

Observation 2.6.8. *In $\mathcal{A}_2^{\leq 4}$ there are no non-trivial colourability sinks or sources.*

Theorem 2.6.9. *For $n \in \mathbb{N}$ let S_n be the number of colourability sinks in $\mathcal{A}_2^n$. If $n \geq 3$,*

$$S_n = K_n - 4.$$

Proof. Let $n \geq 3$ and $w \in \mathcal{A}_2^n$ be a colourability sink. Then by lemma 2.1.10 $\xi(r(w)) = 0$. First, let n be odd. By lemma 2.1.13

$$\xi(w\mathbf{0}) = \xi(w\mathbf{1}) = 0 \iff \xi(r(w)) = 0 \wedge w \notin \left\{\mathbf{0}^n, \mathbf{01}^{n-1}, \mathbf{10}^{n-1}, \mathbf{1}^n\right\},$$

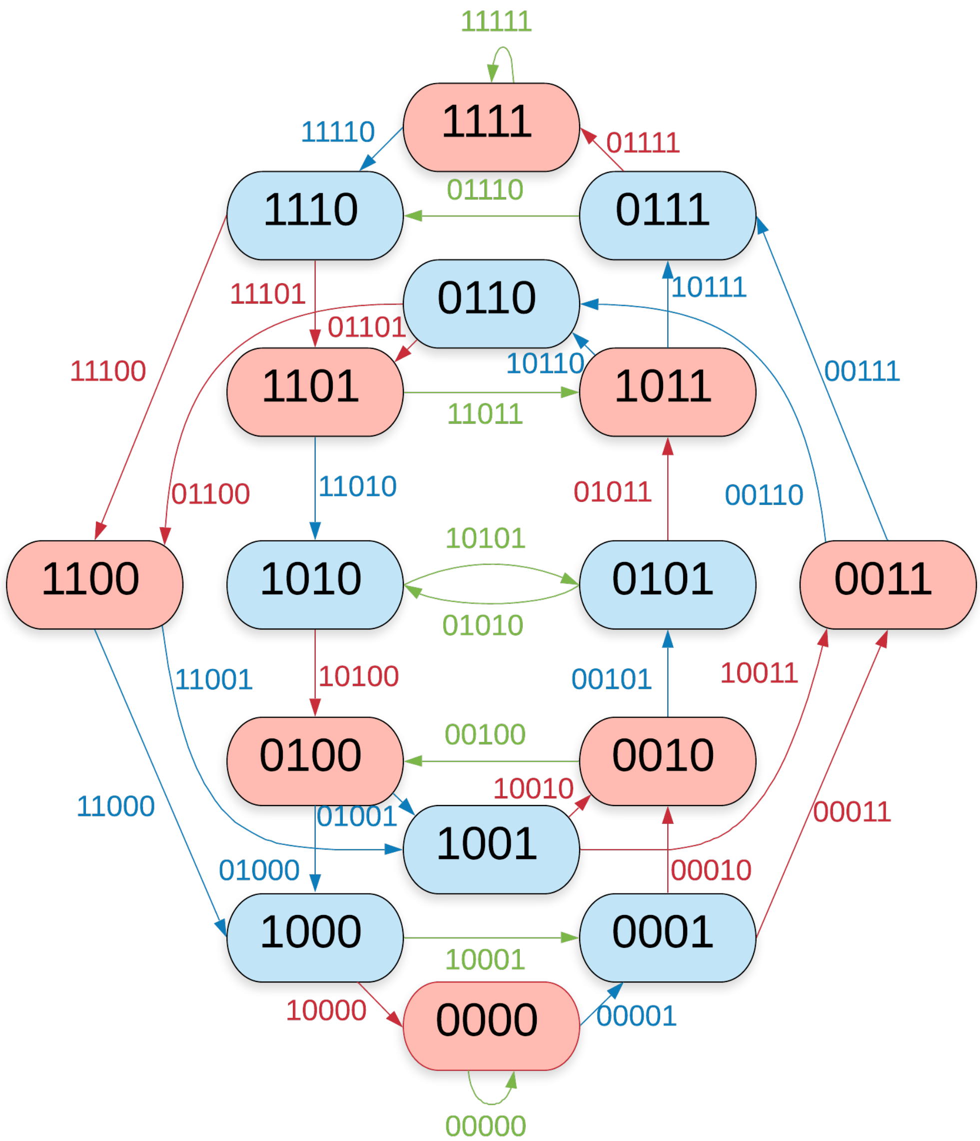

Figure 2-17: The 4-dimensional de Bruijn graph. As in figure 2-10 the edges are coloured according to the value ψ assigns to them. The colour of the vertices coincides with the colour of the incoming edges and differs from the colour of the outgoing ones.

n	0	1	2	3	4	5	6	7	8	9	10	11
S_n	1	0	0	0	0	4	8	20	40	84	168	340

Table 2-18: The number S_n of colourability sinks in $\mathcal{A}_2^n$. There are equally many colourability sources as sinks.

---

so

$$S_n = \#\Big\{w \in \mathcal{A}_2^n; \xi\big(r(w)\big) = 0\Big\} - 4 = 2 \cdot K_{n-1} - 4 \overset{(2.5\text{-}8)}{=} K_n - 4.$$

Now let n be even. Then by lemma 2.1.13

$$\begin{aligned} &\xi\big(w\mathbf{0}\big) = \xi\big(w\mathbf{1}\big) = 0 \\ \iff &\xi\big(r(w)\big) = 0 \wedge w \notin \Big\{\mathbf{0}^n, \mathbf{0}T^{n-1}, T^n, \mathbf{01}^{n-1}, \mathbf{10}^{n-1}, CT^n, \mathbf{1}CT^{n-1}, \mathbf{1}^n\Big\}, \end{aligned}$$

so

$$\begin{aligned} S_n &= \#\Big\{w \in \mathcal{A}_2^n; \xi\big(r(w)\big) = 0\Big\} - 8 \\ &= 2 \cdot K_{n-1} - 8 \\ &\overset{(2.5\text{-}7)}{=} 2 \cdot \frac{2^{n-1} + 16}{6} - 8 \\ &= \frac{2^n + 8}{6} - 4 \\ &\overset{(2.5\text{-}6)}{=} K_n - 4. \end{aligned}$$

Remark. Theorem 2.6.9 is a little bit more surprising than it looks at first glance: Due to the definition of a colourability sink focusing on $w\mathbf{0}$ and $w\mathbf{1}$, S_n rather says something about $\mathcal{A}_2^{n+1}$ than about $\mathcal{A}_2^n$. K_n in contrast should be seen as an information about $\mathcal{A}_2^n$.

A Table with ξ, ϕ and ψ

w	$\xi(w)$	$\phi(w)$	$\psi(w)$
ϵ	0	0	0
0	0	-1	0
1	0	1	0
00	0	0	0
01	1	1	1
10	-1	-1	-1
11	0	0	0
000	0	-1	0
001	1	1	1
010	0	-1	0
011	1	-1	-1
100	-1	1	-1
101	0	1	0
110	-1	-1	1
111	0	1	0
0000	0	0	0
0001	1	1	1
0010	1	-1	-1
0011	1	-1	-1
0100	-1	1	1
0101	1	1	1
0110	0	0	0
0111	1	1	1
1000	-1	-1	-1
1001	0	0	0
1010	-1	-1	-1
1011	1	-1	-1
1100	-1	1	1
1101	-1	1	1

w	$\xi(w)$	$\phi(w)$	$\psi(w)$
1110	-1	-1	-1
1111	0	0	0
00000	0	-1	0
00001	1	1	1
00010	1	-1	-1
00011	1	-1	-1
00100	0	1	0
00101	1	1	1
00110	1	1	1
00111	1	1	1
01000	-1	-1	1
01001	-1	-1	1
01010	0	-1	0
01011	1	-1	-1
01100	-1	1	-1
01101	-1	1	-1
01110	0	-1	0
01111	1	-1	-1
10000	-1	1	-1
10001	0	1	0
10010	1	-1	-1
10011	1	-1	-1
10100	-1	1	-1
10101	0	1	0
10110	1	1	1
10111	1	1	1
11000	-1	-1	1
11001	-1	-1	1
11010	-1	-1	1
11011	0	-1	0

w	$\xi(w)$	$\phi(w)$	$\psi(w)$
11100	-1	1	-1
11101	-1	1	-1
11110	-1	-1	1
11111	0	1	0
000000	0	0	0
000001	1	1	1
000010	1	-1	-1
000011	1	-1	-1
000100	1	1	1
000101	1	1	1
000110	1	1	1
000111	1	1	1
001000	-1	-1	-1
001001	-1	-1	-1
001010	1	-1	-1
001011	1	-1	-1
001100	0	0	0
001101	0	0	0
001110	1	-1	-1
001111	1	-1	-1
010000	-1	1	1
010001	-1	1	1
010010	0	0	0
010011	0	0	0
010100	-1	1	1
010101	1	1	1
010110	1	1	1
010111	1	1	1
011000	-1	-1	-1
011001	-1	-1	-1
011010	-1	-1	-1
011011	-1	-1	-1
011100	-1	1	1
011101	-1	1	1
011110	0	0	0
011111	1	1	1

w	$\xi(w)$	$\phi(w)$	$\psi(w)$
100000	-1	-1	-1
100001	0	0	0
100010	1	-1	-1
100011	1	-1	-1
100100	1	1	1
100101	1	1	1
100110	1	1	1
100111	1	1	1
101000	-1	-1	-1
101001	-1	-1	-1
101010	-1	-1	-1
101011	1	-1	-1
101100	0	0	0
101101	0	0	0
101110	1	-1	-1
101111	1	-1	-1
110000	-1	1	1
110001	-1	1	1
110010	0	0	0
110011	0	0	0
110100	-1	1	1
110101	-1	1	1
110110	1	1	1
110111	1	1	1
111000	-1	-1	-1
111001	-1	-1	-1
111010	-1	-1	-1
111011	-1	-1	-1
111100	-1	1	1
111101	-1	1	1
111110	-1	-1	-1
111111	0	0	0
0000000	0	-1	0
0000001	1	1	1
0000010	1	-1	-1
0000011	1	-1	-1

w	$\xi(w)$	$\phi(w)$	$\psi(w)$
0000100	1	1	1
0000101	1	1	1
0000110	1	1	1
0000111	1	1	1
0001000	0	-1	0
0001001	0	-1	0
0001010	1	-1	-1
0001011	1	-1	-1
0001100	1	-1	-1
0001101	1	-1	-1
0001110	1	-1	-1
0001111	1	-1	-1
0010000	-1	1	-1
0010001	-1	1	-1
0010010	-1	1	-1
0010011	-1	1	-1
0010100	0	1	0
0010101	1	1	1
0010110	1	1	1
0010111	1	1	1
0011000	-1	-1	1
0011001	-1	-1	1
0011010	-1	-1	1
0011011	-1	-1	1
0011100	0	1	0
0011101	0	1	0
0011110	1	1	1
0011111	1	1	1
0100000	-1	-1	1
0100001	-1	-1	1
0100010	0	-1	0
0100011	0	-1	0
0100100	1	1	1
0100101	1	1	1
0100110	1	1	1
0100111	1	1	1

w	$\xi(w)$	$\phi(w)$	$\psi(w)$
0101000	-1	-1	1
0101001	-1	-1	1
0101010	0	-1	0
0101011	1	-1	-1
0101100	1	-1	-1
0101101	1	-1	-1
0101110	1	-1	-1
0101111	1	-1	-1
0110000	-1	1	-1
0110001	-1	1	-1
0110010	-1	1	-1
0110011	-1	1	-1
0110100	-1	1	-1
0110101	-1	1	-1
0110110	0	1	0
0110111	0	1	0
0111000	-1	-1	1
0111001	-1	-1	1
0111010	-1	-1	1
0111011	-1	-1	1
0111100	-1	1	-1
0111101	-1	1	-1
0111110	0	-1	0
0111111	1	-1	-1
1000000	-1	1	-1
1000001	0	1	0
1000010	1	-1	-1
1000011	1	-1	-1
1000100	1	1	1
1000101	1	1	1
1000110	1	1	1
1000111	1	1	1
1001000	0	-1	0
1001001	0	-1	0
1001010	1	-1	-1
1001011	1	-1	-1

w	$\xi(w)$	$\phi(w)$	$\psi(w)$
1001100	1	-1	-1
1001101	1	-1	-1
1001110	1	-1	-1
1001111	1	-1	-1
1010000	-1	1	-1
1010001	-1	1	-1
1010010	-1	1	-1
1010011	-1	1	-1
1010100	-1	1	-1
1010101	0	1	0
1010110	1	1	1
1010111	1	1	1
1011000	-1	-1	1
1011001	-1	-1	1
1011010	-1	-1	1
1011011	-1	-1	1
1011100	0	1	0
1011101	0	1	0
1011110	1	1	1
1011111	1	1	1
1100000	-1	-1	1
1100001	-1	-1	1
1100010	0	-1	0
1100011	0	-1	0
1100100	1	1	1
1100101	1	1	1
1100110	1	1	1
1100111	1	1	1
1101000	-1	-1	1
1101001	-1	-1	1
1101010	-1	-1	1
1101011	0	-1	0
1101100	1	-1	-1
1101101	1	-1	-1
1101110	1	-1	-1
1101111	1	-1	-1
1110000	-1	1	-1
1110001	-1	1	-1
1110010	-1	1	-1
1110011	-1	1	-1
1110100	-1	1	-1
1110101	-1	1	-1
1110110	0	1	0
1110111	0	1	0
1111000	-1	-1	1
1111001	-1	-1	1
1111010	-1	-1	1
1111011	-1	-1	1
1111100	-1	1	-1
1111101	-1	1	-1
1111110	-1	-1	1
1111111	0	1	0

These values have been computed using Python.

B Popular scientific summaries

English

Imagine an *infinite string* of characters A and B. Such can be thought of as a paper strip with As and Bs written on it, one end of which one holds in the hand while the other stretches over the horizon without ever ending. Or as a calculator that displays a new character each second. Or as a printer with an infinite amount of paper and ink that unfortunately only prints As and Bs. In contrast to infinite strings there are also *finite strings* that have a finite length. An infinite string could for instance start like this:

– *infinite string*

– *finite strings*

ABBAABAAABBAABAA...

At this stage any combination (of As and Bs) is possible. The variety shall now be shrunk: For example one can forbid the finite string AA. That means there can never be an A that follows directly on another A. The infinite string consists then of mostly Bs that are just sometimes interrupted by single As:

ABBABABBABABBBAB...

Instead one could also forbid the string AB. Although even here just a string of length 2 is forbidden the effect is considerably larger: As soon as there is an A in the infinite string the As have to continue forever because there will never be a possibility to switch back to a B. The only infinite strings still permitted would be

- the infinite string that only contains As,
- the infinite string that only contains Bs and
- an infinite string that starts with a certain number of Bs and continues with infinitely many As:

BBBAAAAAAAAAAAAA...

By forbidding longer words and several at the same time on can influence the infinite string more subtly: Forbidding the words AAA and ABA for example almost does not have consequences for the appearance of the infinite string. All strings of length 2 are still permitted and can occur in arbitrary order in the infinite string:

ABBAABBBABBABBBA...

Not so if one forbids the word BAA: The string AA can still occur in the beginning of the infinite string but then never again:

AABBABABBABABBBA...

By forbidding strings of a certain length it is hence possible to even forbid shorter strings in practice. Such is called *disconnecting* the infinite string. Using that terminology what has been previously explained can now be put the following way: 'Forbidding AB or BAA disconnects the infinite string, forbidding AA or AAA and ABA does not.' *– disconnect*

Chapter 2 Now each finite string gets a colour. As an example the strings of length 3 shall be considered. The first letter of the strings AAA and ABA are coloured red; the first letters of the words AAB, BAA and BAB blue. An infinite string could start like this:

AAABABAAABAAABAB...

Each character is coloured according to the string of length 3 whose first letter it is: The first string of length 3 is AAA so the first character is red. The second string is AAB so the second character is blue. The last two characters cannot be coloured yet because their colour would depend on the two characters succeeding them.

That red and blue characters alternate is not a coincidence but the very intent with this colouring. The question chapter 2 investigates is: Which finite strings should be coloured red and which ones blue to get such an alternating pattern in the infinite string?

For many infinite strings that is just impossible. The first two characters of an infinite word starting with

AAAA...

will always have the same colour. The solution is here to forbid the string AAAA. Analogously of course one also has to forbid the word BBBB. It turns out that that is not enough either: One also has to forbid the words ABBA and BAAB to find a colouring. It is outlined in table B-0.

A coloured infinite string could then look like this:

AAABABBBAAABBBAB...

Note that two words of length 3 that have an even number of characters between them must always have different colours. Conversely if there is an odd number of characters between them, they have the same colour. If a finite string occurs twice in the infinite string there has to be an odd number of characters between them. Otherwise the same finite string would first have one colour and later another.

Coloured red	Coloured blue	Forbidden
AAA	AAB	AAAA
ABA	ABB	ABBA
BAA	BAB	BAAB
BBA	BBB	BBBB

Table B-0: Possible colouring for strings of length 3 with forbidden strings of length 4

Chapter 1 Forbidding AAAA, ABBA, BAAB and BBBB does not disconnect the infinite string; all strings of length 3 are still permitted and so they are in arbitrary order. They are however bound to certain positions. What is unusual, the restrictions are valid no matter how far one goes in the infinite string.

Forbidding some finite strings often leads to a certain minimum distance between those that are still allowed. If one forbids the string AAA for example there must always be a B between one AA and the next. After that minimum distance however the string AA can again occur at any point. That is not the case when forbidding AAAA, ABBA, BAAB and BBBB: Even if one goes 100 steps further, 1000 or a million: Right there an AAA won't be allowed, just because 100, 1000 and a million are even numbers and between an AAA and the next there must be an odd number of characters.

Constrains that structure the infinite string forever are said to *prevent mixing*.

– prevent mixing

How many ways are there to prevent mixing without disconnecting the infinite string? The answer is given in chapter 1: There aren't that many. If the infinite string does not get disconnected in less than 1 out of 16 cases mixing is prevented.

Swedish – Populärvetenskaplig sammanfattning

En oändlig lång kedja av A:n och B:n får man föreställa sig som en papperslapp där det står A:n och B:n på, dess ena sida man håller i handen men som sträcker sig över horisonten utan att sluta någonstans. Eller som en miniräknare som visar ett nytt tecken varje sekund. Eller en skrivare med oändlig tillgång till papper och bläck men som tyvärr bara skriver ut A:n och B:n. En sådan oändlig lång kedja kallas för *bokstavsföljd* en som har ändlig längd kallas för *ord*. En bokstavsföljd skulle kunna börja så här:

– bokstavsföljd

ABBAABAAABBAABAA...

– ord

Än så länge är varje bokstavskombination (av A:n och B:n) möjlig. Urvalet ska nu begränsas: Man kan till exempel förbjuda ordet AA. Det betyder att det aldrig kan förekomma två A:n direkt efter varandra. Bokstavsföljden består då av en massa B:n som bara ibland bryts av enskilda A:n:

ABBABABBABABBBAB...

Istället skulle man också kunna förbjuda ordet AB. Trots att det igen förbjuds ett ord av längd 2 är effekten betydligt större: Så snart det förekommer ett A i bokstavsföljden kommer A:na fortsätta i all evighet för det finns aldrig en möjlighet att byta tillbaka till ett B. De enda bokstavsföljder som fortfarande är tillåtna är

- bokstavsföljden som bara innehåller A:n,
- bokstavsföljden som bara innehåller B:n och
- en sådan som börjar med ett visst antal B:n och fortsätter med oändligt många A:n:

BBBAAAAAAAAAAAAA...

Genom att förbjuda längre ord och flera samtidigt kan man påverka bokstavsföljden lite mer subtilt: Att förbjuda orden AAA och ABA till exempel har nästan inga konsekvenser för hur bokstavsföljden kommer att se ut. Alla ord med längd 2 är fortfarande tillåtna och kan förekomma i valfri ordningsföljd i bokstavsföljden:

ABBAABBBABBABBBA...

Annars är det när man förbjuder ordet BAA: Ordet AA kan då fortfarande förekomma i början av bokstavsföljden, sedan dock aldrig igen:

AABBABABBABABBBA...

När man förbjuder ord av en viss längd kan man genom detta i praktiken alltså även förbjuda kortare ord. Det kallas för att man *bryter* bokstavsföljden. Nu kan man uttrycka det som sades innan på följande vis: ”Att förbjuda AB eller BAA bryter bokstavsföljden, att förbjuda AA eller AAA och ABA gör inte det.” – *bryta*

Kapitel 2 Nu får varje ord en färg. Som exempel ska ord av längd 3 betraktas. Initialerna av orden AAA och ABA färgas röda; initialerna av orden AAB, BAA och BAB blåa. En bokstavsföljd skulle kunna börja såhär:

AAABABAAABAAABAB...

Varje bokstav är färgad enligt ordet av längd 3 vars initial den är: Första ordet av längd 3 är AAA så första bokstaven är röd. Andra ordet är AAB så andra bokstaven är blå. De sista två bokstäverna går inte att färgas eftersom färgen beror alltid på de två bokstäverna som följer efter.

Att röda och blåa bokstäver alternerar är inget slump utan målet med färgningen. Frågan som kapitel 2 tittar på är: Vilka ord borde man färga röda och vilka blåa för att få ett sådant alternerande mönster i bokstavsföljden?

För många bokstavsföljder går det över huvud taget inte. In en sådan som börjar med

AAAA...

kommer de första två bokstäverna alltid ha samma färg. Lösningen är här att förbjuda ordet AAAA. Likaså får man såklart förbjuda ordet BBBB. Det visar sig att inte ens det räcker: Man behöver även förbjuda orden ABBA och BAAB för att hitta ett sätt. Det beskrivs i tabell B-1.

Röda färgade	Blåa färgade	Förbjudna
AAA	AAB	AAAA
ABA	ABB	ABBA
BAA	BAB	BAAB
BBA	BBB	BBBB

Tabell B-1: Möjlig färgning till orden av längd 3 med förbud av vissa ord av längd 4

En färgad bokstavsföljd skulle då kunna börja såhär:

AAABABBBAAABBBAB...

Man inser: Två ord av längd 3 som har ett jämnt antal bokstäver emellan sig måste alltid har olika färger. Står däremot ett udda tal bokstäver emellan dem så har de samma färg. När ett ord förekommer två gånger i bokstavsföljden måste det alltså finnas ett udda tal bokstäver mellan dem. Annars skulle samma ord först ha en färg och sedan en annan.

Kapitel 1 Att samtidigt förbjuda orden AAAA, ABBA, BAAB och BBBB bryter inte bokstavsföljden; alla ord av längd 3 är fortfarande tillåtna och det i valfri ordning. De är dock bundna till vissa positioner. Ovanligt är att begränsningen kvarstår oavsett hur långt man går i bokstavsföljden.

Att förbjuda några ord leder ofta till en viss minimiavsstånd mellan de orden som fortfarande är tillåtna. Förbjuder man ordet AAA till exempel måste det alltid finnas ett B mellan ett AA och det nästa. Efter minimiavståndet kan ordet dock förekommer igen när som helst. Det gäller inte när man förbjuder AAAA, ABBA, BAAB och BBBB: Även om man går 100 steg vidare, 1000 eller en miljon: Där får inte stå ett AAA, helt enkelt för 100, 1000 och en miljon är jämna tal och mellan ett AAA och ett annat så måste det alltid finnas ett udda antal bokstäver.

Inskränkningar som strukturerar bokstavsföljden för alltid sägs *förhindra blandning*.

– förhindra blandning

Hur många sätt finns det att förhindra blandning utan att bryta bokstavsföljden? Svaret ges i kapitel 1: Finns inte särskilt många. Om man inte bryter bokstavsföljden så förhindras blandning i mindre än 1 av 16 fall.

German – Populärwissenschaftliche Zusammenfassung

Man denke sich eine unendlich lange Aneinanderreihung der Buchstaben A und B. Das kann man sich vorstellen wie einen Papierstreifen auf dem As und Bs stehen, dessen Anfang man in der Hand hält, der sich aber über den Horizont erstreckt und nie endet. Oder wie einen Taschenrechner, der jede Sekunde ein neues Zeichen anzeigt. Oder einen Drucker mit unendlichem Vorrat an Papier und Toner, der aber leider nur As und Bs druckt. Ein solche unendlich lange Aneinanderreihung wird *Buchstabenfolge* genannt, eine endlich lange dagegen ein *Wort*. Eine Buchstabenfolge könnte zum Beispiel folgendermaßen beginnen:

– Buchstabenfolge
– Wort

ABBAABAAABBAABAA...

Bisher ist jede beliebe Buchstabenkombination (von As und Bs) möglich. Die Auswahl soll jetzt eingeschränkt werden: Man kann zum Beispiel das Wort AA verbieten. Das bedeutet, dass nie ein A auf ein anderes folgen kann. Die Buchstabenfolge besteht aus einem Haufen Bs, nur manchmal von einzelnen As unterbrochen:

ABBABABBABABBBAB...

Stattdessen könnte man auch das Wort AB ausschließen. Obwohl auch hier nur ein Wort der Länge 2 verboten wird, sind die Auswirkungen viel weitreichender: Sobald ein A in der Buchstabenfolge auftaucht, werden sich As bis in alle Ewigkeit fortziehen, da es nie die Möglichkeit gibt, wieder zu einem B zurückzukehren. Die einzigen Buchstabenfolgen, die jetzt noch möglich sind, sind

- die Buchstabenfolge, die ausschließlich aus As besteht,
- die Buchstabenfolge, die ausschließlich aus Bs besteht und
- eine solche, die mit einer bestimmten Anzahl Bs anfängt und sich dann mit unendlich vielen As fortsetzt:

BBBAAAAAAAAAAAAA...

Indem man längere Wörter verbietet, und mehrere gleichzeitig, kann man die Buchstabenfolge subtiler beeinflussen: Das Verbieten der Wörter AAA und ABA zum Beispiel hat kaum Auswirkungen auf das Aussehen der Buchstabenfolge. Alle Wörter der Länge 2 sind noch erlaubt und können in beliebiger Reihenfolge in der Buchstabenfolge vorkommen:

ABBAABBBABBABBBA...

Anders sieht das aus, wenn man das Wort BAA verbietet: Das Wort AA kann dann zwar ganz am Anfang der Buchstabenfolge vorkommen, dann aber nie wieder:

AABBABABBABABBBA...

Wenn man Wörter einer bestimmten Länge verbietet, kann man damit also praktisch auch kürzere Wörter verbieten. In dem Fall wird gesagt, man *zerreißt* die Buchstabenfolge. Mit diesem Vokabular lässt sich das bisher gesagte folgendermaßen ausdrücken: „Das Verbieten von AB oder BAA zerreißt die Buchstabenfolge, das Verbieten von AA oder AAA und ABA dagegen nicht.“ *– Zerreißen*

Kapitel 2 Jedem Wort wird jetzt eine Farbe zugewiesen. Beispielhaft werden hier Wörter der Länge 3 betrachtet. Die Anfangsbuchstaben der Wörter AAA und ABA werden rot eingefärbt; die Anfangsbuchstaben der Wörter AAB, BAA und BAB blau. Eine Buchstabenfolge könnte dann folgendermaßen beginnen:

AAABABAAABAAABAB...

Jeder Buchstabe ist hier gemäß dem Wort der Länge 3 von dem er der Anfangsbuchstabe ist eingefärbt: Das erste Wort der Länge 3 ist AAA, also ist der erste Buchstabe rot. Das zweite Wort ist AAB, also ist der zweite Buchstabe blau. Die letzten beiden Buchstaben lassen sich noch nicht einfärben, da sich die Farbe immer erst aus den beiden darauffolgenden Buchstaben ergibt.

Dass sich hier rote und blaue Buchstaben abwechseln ist kein Zufall, sondern gerade das Ziel dieser Färbung. Die Frage, der Kapitel 2 nachgeht, ist: Welche Wörter sollte man rot und welche blau einfärben, damit man so ein alternierendes Farbmuster in der Buchstabenfolge bekommt?

Für viele Buchstabenfolgen geht das überhaupt nicht. In einer, die mit

AAAA...

beginnt, werden die ersten beiden Buchstaben immer dieselbe Farbe haben. Der Ausweg hier ist, das Wort AAAA zu verbieten. Genauso muss natürlich das Wort BBBB verboten werden. Es zeigt sich, dass auch das nicht reicht: Erst wenn man darüber hinaus die Wörter ABBA und BAAB verbietet, lässt sich eine Lösung finden. Sie ist in Tabelle B-2 dargestellt.

Rot eingefärbt	Blau eingefärbt	Verboten
AAA	AAB	AAAA
ABA	ABB	ABBA
BAA	BAB	BAAB
BBA	BBB	BBBB

Tabelle B-2: Mögliche Färbung der Wörter der Länge 3 unter Ausschluss bestimmter Wörter der Länge 4

Eine eingefärbte Buchstabenfolge könnte dann folgendermaßen beginnen:

AAABABBBAAABBBAB...

Man kann erkennen: Zwei Wörter der Länge 3, zwischen denen eine gerade Zahl von Buchstaben steht, müssen immer verschiedene Farben haben. Steht dagegen eine ungerade Zahl von Buchstaben zwischen ihnen, haben sie die gleiche Farbe. Wenn ein Wort zweimal in der Buchstabenfolge vorkommt, muss also auch eine ungerade Zahl an Buchstaben dazwischen stehen. Ansonsten hätte dasselbe Wort erst die eine und später eine andere Farbe.

Kapitel 1 Das gleichzeitige Verbieten der Wörter AAAA, ABBA, BAAB und BBBB zerreißt die Buchstabenfolge nicht; alle Wörter der Länge 3 sind weiterhin erlaubt und zwar in beliebiger Reihenfolge. Sie sind aber an bestimmte Positionen gebunden. Ungewöhnlich dabei ist, dass Begrenzungen bestehen bleiben, unabhängig davon, wie weit man in der Buchstabenfolge geht.

Das Verbieten einiger Wörter führt häufig zu einem gewissen Mindestabstand zwischen den Wörtern, die weiterhin erlaubt sind. Verbietet man das Wort AAA zum Beispiel muss mindestens ein B zwischen einem AA und dem nächsten AA stehen. Nach dem Mindestabstand aber kann das Wort an beliebiger Stelle wieder auftreten. Das ist bei dem Verbieten von AAAA, ABBA, BAAB und BBBB anders: Selbst wenn man nach einem Auftreten des Wortes AAA 100 Schritte weiter geht, 1000 oder eine Million: An der Stelle darf kein AAA stehen, schlicht weil 100, 1000 und eine Million gerade Zahlen sind und zwischen einem AAA und dem nächsten immer eine ungerade Zahl an Buchstaben stehen muss.

Vorgaben, die die Buchstabenfolge bis auf alle Ewigkeit strukturieren werden *Durchmischung verhindernd* genannt.

– Durchmischung verhindern

Wie viele Möglichkeiten gibt es, Durchmischung zu verhindern, ohne die Buchstabenfolge zu zerreißen? Die Antwort wird in Kapitel 1 gegeben: Nicht sehr viele. Wenn man die Buchstabenfolge nicht zerreißt, wird in weniger als einem von 16 Fällen Durchmischung verhindert.

Bibliography

de Bruijn, Nicolaas Govert (1946). 'A combinatorial problem'. English. In: *Proceedings of the Section of Sciences of the Koninklijke Nederlandse Akademie van Wetenschappen te Amsterdam* 49.7, pp. 758–764.

– (1975). *Acknowledgement of priority to C. Flye Sainte-Marie on the counting of circular arrangements of 2^n zeros and ones that show each n-letter word exactly once*. English. Eindhoven.

Flye Sainte-Marie, Camille (1894). 'Solution to question nr. 48'. In: *L'Intermédiaire des Mathématiciens* 1, pp. 107–110.

Lind, Douglas and Brian Marcus (1995). *An introduction to symbolic dynamics and coding*. English. Cambridge: Cambridge University Press. ISBN: 0-521-55124-2.

Wikipedia contributors (2020a). *De Bruijn graph*. Wikipedia, The Free Encyclopedia. URL: https://en.wikipedia.org/w/index.php?title=De_Bruijn_graph&oldid=959972670 (visited on 2020-06-03).

– (2020b). *Glossary of graph theory terms*. Wikipedia, The Free Encyclopedia. URL: https://en.wikipedia.org/w/index.php?title=Glossary_of_graph_theory_terms&oldid=946797667 (visited on 2020-05-23).

– (2020c). *Jacobsthal number*. Wikipedia, The Free Encyclopedia. URL: https://en.wikipedia.org/w/index.php?title=Jacobsthal_number&oldid=954813000 (visited on 2020-06-03).