Phased Multi de Bruijn sequences

Glenn Tesler

Submitted: Jul 23, 2022; Accepted: Nov 25, 2025; Published: Apr 14, 2026

Abstract

We introduce *phased multi de Bruijn sequences*, a generalization of de Bruijn sequences. A *phased string* is a string whose positions sequentially rotate through several alphabets; e.g., "$0Ax1Ay1By0Az1Bx$" rotates through alphabets $\Omega_0 = \{0, 1\}$, $\Omega_1 = \{A, B\}$, and $\Omega_2 = \{x, y, z\}$. We consider cyclic phased strings in which all possible phased k-mers (phased strings of length k) occur with particular multiplicities, depending on their "phase" (the alphabet they start in). For example, consider the cycle $(s) = (0Ax1Ay1By0Az0By1Bx0Ay0Bx1Bz1Ax0Bz1Az)$ in these alphabets. All possible phased 2-mers starting in phases 0, 1, and 2 respectively have multiplicities 3, 2, and 2; e.g., "$0A$" occurs three times, "Ax" occurs twice, and "$z0$" occurs twice (including the occurrence that wraps around the cycle). We determine parameters (k, number of phases, alphabet sizes, and multiplicities) for which this is possible. Then we count the total number of phased multi de Bruijn sequences for these parameters. This extends classical de Bruijn sequences and multi de Bruijn sequences (our previous generalization of de Bruijn sequences in which all possible k-mers over one alphabet occur m times each). Our method of counting the sequences uses a change of basis for the Laplacian matrix; this also gives a new proof for the number of classical de Bruijn sequences, as they are a special case of this framework.

Mathematics Subject Classifications: 05C30, 05C38, 05C45, 05A15

1 Introduction

We present *phased multi de Bruijn sequences*, a generalization of de Bruijn sequences in which positions cycle through multiple alphabets. In this introduction, we'll define notation for linear and circular sequences; review de Bruijn sequences and a generalization, multi de Bruijn sequences; and then introduce phased multi de Bruijn sequences.

Department of Mathematics, University of California San Diego, La Jolla, CA, U.S.A. (gtesler@ucsd.edu).

1.1 Notation preliminaries

A *linear sequence* is a sequence $s_0, s_1, \ldots, s_{r-1}$ of characters over a totally ordered alphabet, Ω. We use string notation to denote it as $s = s_0 s_1 \ldots s_{r-1}$.

The *cyclic shift* of this sequence is $\rho(s_0 s_1 \ldots s_{r-2} s_{r-1}) = s_{r-1} s_0 s_1 \ldots s_{r-2}$. Shifting i positions to the right is denoted by $\rho^i(s)$. A *cyclic sequence*, denoted with parentheses as (s), treats all rotations as equivalent: $(s) = \{\rho^i(s) \ : \ i = 0, \ldots, r-1\}$. E.g.,

$$\begin{aligned}(001001) &= \{001001, 100100, 010010, 001001, 100100, 010010\} \\ &= \{001001, 100100, 010010\}.\end{aligned}$$

Each rotation $\rho^i(s)$ is called a *linearization* of the cycle (s); here, (001001) has three distinct linearizations, since some of the six rotations are identical.

A *linearized sequence* is a linear sequence used to represent a cyclic sequence: its positions are fixed, but we wrap around from the last character to the first character.

For an integer $k \geqslant 0$, a *k-mer* is a sequence of length k over Ω. A linear sequence of length r has a k-mer starting at each of the first $r - k + 1$ positions. For *linearized sequences*, we also consider k-mers that "wrap around" the sequence, thus giving a k-mer starting at every position. In linearized sequence $s = s_0 \ldots s_5 = 001001$, the 3-mer starting at position 5 is $s_5 s_0 s_1 = 100$. If $k > r$, wrap around s as many times as needed: the 8-mer starting at position 5 is $s_5 s_0 s_1 s_2 s_3 s_4 s_5 s_0 = 10010010$.

For a linear sequence s and integer $h \geqslant 0$, let s^h be the concatenation of h copies of s. E.g., $s = 001$ gives $s^2 = 001001$.

A sequence s or (s) of length r has a d^{th} *order rotation* iff $\rho^{r/d}(s) = s$; or equivalently, iff $s = t^d$ for some sequence t of length r/d. This requires that d be a positive divisor of r (denoted $d|r$). The *rotational order* of (s) or s is the largest $d|r$ such that $\rho^{r/d}(s) = s$; e.g., (001001) has order 2.

1.2 De Bruijn sequences and multi de Bruijn sequences

Let $q, k \geqslant 1$. Consider cyclic sequences over an alphabet of size q. A *de Bruijn sequence* is a cyclic sequence in which all q^k possible k-mers occur exactly once. For example, with alphabet $\{0, 1\}$ of size $q = 2$, and words of length $k = 2$, the cyclic string (0011) has all four 2-mers (00, 01, 10, 11) exactly once; this includes an occurrence of 10 that wraps around the cyclic sequence from the end to the start. In 1894, de Rivière [7] proposed counting such sequences for $q = 2$ and any k, and Sainte-Marie [15] solved it, but this work was nearly lost to history. The same case ($q = 2$, $k \geqslant 1$) was independently solved in 1946 by de Bruijn [5], and extended in 1951 to all $q, k \geqslant 1$ by van Aardenne-Ehrenfest and de Bruijn [20, p. 203]. In 1975, Richard Stanley rediscovered the 1894 work by de Rivière and Sainte-Marie; see [6] for further background.

The solutions [15, 5, 20] use a graph construction and induct on k (with q fixed). In addition, van Aardenne-Ehrenfest and de Bruijn [20] also proved what is now called the BEST Theorem, which counts the number of cycles in a graph using a determinant, but they did not actually evaluate the determinant. In 1957, Dawson and Good [4] were the first to evaluate the determinant to count de Bruijn sequences for arbitrary q and k; they noted that a power of the graph's transition matrix had constant columns, leading to a simple analysis of its eigenvalues.

Let $m \geqslant 1$. A *multi de Bruijn sequence* is a cyclic sequence with each of the q^k possible k-mers occurring exactly m times. For example, (011101000110) has $m = 3$ instances of each of 00, 01, 10, and 11, including an instance of 00 that wraps around. In [18], I introduced multi de Bruijn sequences and counted them, also using the BEST Theorem and evaluating the determinant by using a matrix power.

Here we develop a further generalization, *phased multi de Bruijn sequences*. We use a change of basis to evaluate the determinant in the BEST Theorem. Since classical de Bruijn sequences are a special case of this framework, this gives a new proof of the formula for the number of de Bruijn sequences. In the new basis, the Laplacian matrix is sparse (at most two nonzero entries per column), and for classical de Bruijn sequences, it is lower triangular, allowing easy evaluation of the determinant. In Sec. 3.2, we will also develop a generalization of Dawson and Good's [4] result about a power of the transition matrix.

A *multicyclic sequence* is a multiset of aperiodic cyclic sequences. Higgins [10] introduced an analog of de Bruijn sequences: a multicyclic sequence containing every k-mer exactly once, and Tesler [18] extended this to multicyclic sequences containing every k-mer exactly m times. E.g., (00111)(001) contains each of 00, 01, 10, and 11 exactly twice. In future work in preparation, we will extend this to multicyclic analogs of other universal cycle problems, including a multicyclic analog of phased multi de Bruijn sequences.

1.3 Phased multi de Bruijn sequences

Example 1. Consider this cyclic string (one string, written on two lines due to length):

$$\begin{array}{l}(0A0A0B0B1A0C0B0C0C1C1B0A0B1C0B0A1A0C\\ \ 0A1B1A1C1C0A1C1A1A1B1B0B1C0C1B1B0C1A)\end{array} \tag{1}$$

The length is 72. Number the positions $0, 1, \dots, 71$ from left to right. Characters in even numbered positions are from the alphabet $\Omega_0 = \{0, 1\}$, while odd numbered positions are from $\Omega_1 = \{A, B, C\}$. The 3-mers starting at even numbered positions, such as $0A0$ at the start, are in *phase* 0, while the 3-mers starting at odd numbered positions, such as $A0A$, are in *phase* 1. Phase 0 has $2 \cdot 3 \cdot 2 = 12$ possible 3-mers (see Table 1), each occurring three times in the sequence above, while phase 1 has $3 \cdot 2 \cdot 3 = 18$ possible 3-mers, each occurring twice. We will determine alphabet sizes and multiplicities for which sequences like this are possible, and count how many such sequences there are.

We let $p \in \mathbb{Z}^+$ be the number of *phases*; note that p need not be prime. We denote the set of phases by $\mathbb{P} = \mathbb{Z}/p\mathbb{Z} = \{0, 1, \dots, p-1\}$, with addition modulo p.

Form p disjoint alphabets Ω_i, for $i \in \mathbb{P}$. Set $q_i = |\Omega_i|$ and $\vec{q} = (q_0, \dots, q_{p-1})$. This example has two phases ($p = 2$) and alphabet sizes $q_0 = 2$ and $q_1 = 3$, giving $\vec{q} = (2, 3)$. Since phases are numbered modulo $p = 2$, we have $q_0 = q_2 = \cdots = 2$; $q_1 = q_3 = \cdots = 3$; $\Omega_0 = \Omega_2 = \cdots = \{0, 1\}$; and $\Omega_1 = \Omega_3 = \cdots = \{A, B, C\}$.

Table 1: Sets of words for Example 1. Alphabets $\Omega_0 = \{0, 1\}$, $\Omega_1 = \{A, B, C\}$; alphabet sizes $q_0 = 2$, $q_1 = 3$; and word size $k = 3$.

Phase	**Set**	**Set size**
***k*-mers (3-mers)**		
Phase 0	$K_0 = \Omega_0 \times \Omega_1 \times \Omega_2 = \Omega_0 \times \Omega_1 \times \Omega_0$ $= \{0A0, 0A1, 0B0, 0B1, 0C0, 0C1,$ $1A0, 1A1, 1B0, 1B1, 1C0, 1C1\}$	$\theta_0 = q_0\, q_1\, q_2 = q_0\, q_1\, q_0$ $= 2 \cdot 3 \cdot 2 = 12$
Phase 1	$K_1 = \Omega_1 \times \Omega_2 \times \Omega_3 = \Omega_1 \times \Omega_0 \times \Omega_1$ $= \{A0A, A0B, A0C, A1A, A1B, A1C,$ $B0A, B0B, B0C, B1A, B1B, B1C,$ $C0A, C0B, C0C, C1A, C1B, C1C\}$	$\theta_1 = q_1\, q_2\, q_3 = q_1\, q_0\, q_1$ $= 3 \cdot 2 \cdot 3 = 18$
$(k-1)$-mers (2-mers)		
Phase 0	$V_0 = \{0A, 0B, 0C, 1A, 1B, 1C\}$	$n_0 = 6$
Phase 1	$V_1 = \{A0, A1, B0, B1, C0, C1\}$	$n_1 = 6$

Each phase has its own null string (string of length 0), denoted $\emptyset_i$. The set of strings in phase $i \in \mathbb{P}$ with length $r \geqslant 0$ is denoted $\mathcal{S}_{i,r}$, and the set of all phase i strings is denoted $\mathcal{S}_i$. The phase of a string $s \in \mathcal{S}_i$ is denoted by $\textsc{Phase}(s) = i$. We have:

$$\mathcal{S}_{i,0} = \{\emptyset_i\} \qquad \mathcal{S}_{i,r} = \Omega_i \times \Omega_{i+1} \times \cdots \times \Omega_{i+r-1} \quad \text{for } r > 0 \tag{2}$$

$$\mathcal{S}_i = \bigcup_{r=0}^{\infty} \mathcal{S}_{i,r} \ . \tag{3}$$

We write a phased string as a linear sequence $s = s_0 s_1 \dots s_{r-1}$ whose positions cycle through these alphabets: $s_h \in \Omega_{i+h}$ for $h = 0, \dots, r-1$, where $i \in \mathbb{P}$ is the phase of s. With the alphabets of Example 1, $0C0A1B1$ is a phase 0 string of length 7, and $B0C1A0$ is a phase 1 string of length 6. A *phased cyclic sequence* is (s), where s is a phased string with length divisible by p; this condition is required so that adjacent character phases are consecutive, including when wrapping around the cycle. E.g., $(B0C1A0)$ is allowed, but $(0C0A1B1)$ is not: the last and first positions are consecutive in the cycle, but have non-consecutive phases 0 and 0. There are several notation issues:

- Positions in phased string $s = s_0 \dots s_{r-1}$ are indexed by integers, so $s_0, \dots, s_{r-1}$ are separate variables. But Ω_i is indexed by phases (integers mod p); e.g., for $p = 2$, we define Ω_0 and Ω_1 and extend to $\Omega_0 = \Omega_2 = \Omega_4 = \cdots$ and $\Omega_1 = \Omega_3 = \Omega_5 = \cdots$.
- For nonempty strings, the first character sets the phase; e.g., $\textsc{Phase}(B0C1A0) = 1$. For null strings, $\textsc{Phase}(\emptyset_i) = i$ with $i \in \mathbb{P}$.

- For a phased string expressed as a sequence of character variables, the phase should be specified, or implied by the information given. E.g., $s = s_0 \dots s_{r-1} \in \mathcal{S}_i$ indicates that s is in phase i and its characters satisfy $s_h \in \Omega_{i+h}$. This implies substring $s' = s_h\, s_{h+1}\, \dots\, s_{j-1}$ (with $0 \leqslant h \leqslant j \leqslant r$) has $\textsc{Phase}(s') = i + h$.

The number of phase i strings of length r is

$$Q_{i,r} = |\mathcal{S}_{i,r}| = \begin{cases} q_i q_{i+1} \dots q_{i+r-1} & \text{if } r \geqslant 1; \\ |\{\emptyset_i\}| = 1 & \text{if } r = 0. \end{cases} \tag{4}$$

Let $i \in \mathbb{P}$ and $k \geqslant 1$. We define these special cases:

$$\begin{array}{ll} K_i = \mathcal{S}_{i,k} = \text{Set of phase } i \ k\text{-mers} & \text{size: } \theta_i = |K_i| = Q_{i,k} \\ V_i = \mathcal{S}_{i,k-1} = \text{Set of phase } i \ (k-1)\text{-mers} & \text{size: } n_i = |V_i| = Q_{i,k-1} \\ & \text{note: } \theta_i = n_i \cdot q_{i+k-1} = q_i \cdot n_{i+1} \end{array} \tag{5}$$

$$K = K_0 \cup \dots \cup K_{p-1} \qquad\qquad V = V_0 \cup \dots \cup V_{p-1}. \tag{6}$$

See Table 1 for an example. In Sec. 3, we'll form a multigraph on vertices V and edges K.

Let $m_i \in \mathbb{Z}^+$ for $i \in \mathbb{P}$, and set $\vec{m} = (m_0, \dots, m_{p-1})$. A phased cyclic sequence in which every element of K_i occurs exactly m_i times, for each $i \in \mathbb{P}$, is called a *phased multi de Bruijn sequence*. Let $\mathcal{C}(\vec{m}, \vec{q}, k)$ denote the set of all such sequences. We will count the number of such sequences.

Example 1 has $\vec{m} = (3, 2)$, $\vec{q} = (2, 3)$, and $k = 3$. Thus, Eq. (1) is an element of $\mathcal{C}((3, 2), (2, 3), 3)$. Every element of K_0 (see Table 1) occurs in it $m_0 = 3$ times and every element of K_1 occurs $m_1 = 2$ times. The $m_0 = 3$ occurrences of $1A0$ are underlined below, including one that wraps around:

$$\begin{array}{l} (\underline{0}A0A0B0B\underline{1A0}C0B0C0C1C1B0A0B1C0B0A\underline{1A0}C \\ \ 0A1B1A1C1C0A1C1A1A1B1B0B1C0C1B1B0C\underline{1A}) \end{array} \tag{7}$$

Occurrences of a k-mer may overlap; the two overlapping occurrences of $0A0$ in "0A0A0" at the start of the above sequence, count as two distinct occurrences.

Let $\mathcal{LC}(\vec{m}, \vec{q}, k)$ be the set of all linearizations of sequences in $\mathcal{C}(\vec{m}, \vec{q}, k)$, and for $y \in K$, let $\mathcal{LC}_y(\vec{m}, \vec{q}, k)$ be the subset of those that start with y. Eq. (7) contributes 72 linearizations to $\mathcal{LC}((3, 2), (2, 3), 3)$. For $y = 1A0$, Eq. (7) contributes 3 linearizations to $\mathcal{LC}_{1A0}((3, 2), (2, 3), 3)$; each is a length 72 string, written on two lines due to length:

$$\begin{array}{l} \bullet\ 1A0C0B0C0C1C1B0A0B1C0B0A1A0C0A1B1A1C \\ \quad 1C0A1C1A1A1B1B0B1C0C1B1B0C1A0A0A0B0B \\ \bullet\ 1A0C0A1B1A1C1C0A1C1A1A1B1B0B1C0C1B1B \\ \quad 0C1A0A0A0B0B1A0C0B0C0C1C1B0A0B1C0B0A \\ \bullet\ 1A0A0A0B0B1A0C0B0C0C1C1B0A0B1C0B0A1A \\ \quad 0C0A1B1A1C1C0A1C1A1A1B1B0B1C0C1B1B0C \end{array} \tag{8}$$

For $i \in \mathbb{P}$, let $\mathcal{LC}_{[[i]]}(\vec{m}, \vec{q}, k)$ be the set of phase i linearizations of sequences in $\mathcal{C}(\vec{m}, \vec{q}, k)$. Eq. (7) has 36 linearizations in phase 0 (starting with '0' or '1') and 36 linearizations in phase 1 (starting with 'A', 'B', or 'C').

Example 2. We illustrate constraints between $\vec{m}$, $\vec{q}$, and k, by modifying $\vec{m}$ in Example 1. We keep $\vec{q} = (2, 3)$, $k = 3$, and the same alphabets Ω_0, Ω_1, and sets of k-mers K_0, K_1.

If a phased multi de Bruijn sequence with multiplicities $\vec{m} = (m_0, m_1)$ exists, it has $m_0 \cdot |K_0| = 12m_0$ occurrences of elements of K_0 (each starting at a different even numbered position) and $m_1 \cdot |K_1| = 18m_1$ occurrences of elements of K_1 (each starting at a different odd numbered position). The length of the string is even (it's a multiple of $p = 2$), so there is an equal number of even and odd numbered positions: $12m_0 = 18m_1$. Thus, $m_1 = (2/3)m_0$. Since $m_0, m_1 \in \mathbb{Z}^+$, we must have $\vec{m} = (3r, 2r)$ for some $r \in \mathbb{Z}^+$.

To show at least one sequence exists with $\vec{m} = (3r, 2r)$, concatenate r copies of the string in Eq. (1) and circularize it. The multiplicity of every k-mer goes up r-fold. For example, here we double the sequence (s) to obtain $(s^2) \in \mathcal{C}((6, 4), (2, 3), 3)$:

$$\begin{aligned}
&(\underline{0}A0A0B0B\underline{1A0}C0B0C0C1C1B0A0B1C0B0A\underline{1A0}C \\
&\ 0A1B1A1C1C0A1C1A1A1B1B0B1C0C1B1B0C\underline{1A} \\
&\ \underline{0}A0A0B0B\underline{1A0}C0B0C0C1C1B0A0B1C0B0A\underline{1A0}C \\
&\ 0A1B1A1C1C0A1C1A1A1B1B0B1C0C1B1B0C\underline{1A})
\end{aligned} \tag{9}$$

This has $m_0 = 6$ occurrences of $1A0$ (each underlined). But rotating the sequence to start at each of the six occurrences only yields three distinct linearizations instead of six: the first three $1A0$'s give distinct linearizations corresponding to doubling each of the linearizations in (8), while the linearizations starting at the other three $1A0$'s duplicate these. The number of distinct linearizations beginning with a given k-mer depends on the sequence, not just on $\vec{m}$. We will revisit this in Sec. 4.3.

1.4 Outline

Sec. 2 generalizes Example 2 to characterize the parameters $(\vec{m}, \vec{q}, k)$ for which cyclic sequences exist.

Sec. 3 forms the *phased multi de Bruijn graph*, G. We represent phased sequences by walks in G (Sec. 3.1), and determine properties of the adjacency matrix (Sec. 3.2) and vertex degrees (Sec. 3.3).

Sec. 4 counts cyclic phased multi de Bruijn sequences. We state results about the characteristic polynomial of its Laplacian matrix (Sec. 4.1), but defer technical proofs to Sec. 7. Then we use the BEST Theorem to count Eulerian cycles in G (Sec. 4.2) and count cyclic phased multi de Bruijn sequences (Sec. 4.3).

Sec. 5 gives an application of phased multi de Bruijn sequences to a deck of cards, with each card represented by two phases: denominations and suits.

Sec. 6 shows that phased multi de Bruijn sequences are *universal cycles* [2] for certain walks in periodic irreducible Markov chains.

Sec. 7 uses a change of basis to simplify the Laplacian matrix and compute its characteristic polynomial.

In Sec. 8, reversing phased multi de Bruijn sequences gives phased multi de Bruijn sequences with different parameters.

Finally, Appendix A has examples of our notations, matrices, and graphs, for cyclic sequences with one phase (classical and multi de Bruijn sequences) and four phases.

2 Compatible parameters for cyclic sequences

Example 2 illustrated that phased multi de Bruijn sequences only exist for certain parameter values. This section constructs all such parameters.

Assume a phased cyclic multi de Bruijn sequence, (s), exists with parameters $(\vec{m}, \vec{q}, k)$. Let s be a phase 0 linearization. Each of the θ_i elements of K_i occurs exactly m_i times in (s); collectively, these start at $m_i\theta_i$ distinct positions. Conversely, every phase i position of (s) has a phase i k-mer starting at it. Thus, s has $m_i\theta_i$ positions in phase i (namely, positions congruent to i modulo p).

Sequence s repeatedly goes through blocks of positions in consecutive phases $0, 1, \ldots,$ $p-1$, including the last position of s (phase $p-1$) cyclically followed by the first position (phase 0). So s has an equal number of positions in every phase. Thus,

$$m_0\theta_0 = \cdots = m_{p-1}\theta_{p-1} \; . \tag{10}$$

Parameters $(\vec{m}, \vec{q}, k)$ are called *compatible* when k, p, $m_0, \ldots, m_{p-1}$, $q_0, \ldots, q_{p-1}$ are positive integers satisfying (10) (where we expand θ's in terms of q's via (4)–(5)). This is necessary for cyclic sequences to exist, and in Sec. 3.3, we will show it is also sufficient.

The total length of s or (s) is $\ell = \sum_{i=0}^{p-1} m_i\theta_i$; by (10), this simplifies to

$$\ell = p\, m_i\, \theta_i = p\, m_i \; \cdot \; q_i\, q_{i+1} \; \cdots \; q_{i+k-1} \qquad \text{for any } i \in \mathbb{P}. \tag{11}$$

Solve (10) for the multiplicities:

$$m_i = (\theta_0/\theta_i)m_0 \qquad \text{for } i \in \mathbb{P}. \tag{12}$$

Lemma 3. *Let $k \geqslant 0$ and $k' = k \bmod p$. Then $Q_{0,k}/Q_{i,k} = Q_{0,k'}/Q_{i,k'}$.*

Proof. For $0 \leqslant k < p$, we have $k' = k$, so $Q_{0,k}/Q_{i,k} = Q_{0,k'}/Q_{i,k'}$. For $k \geqslant p$, we have

$$Q_{i,k} = \underbrace{q_i q_{i+1} \ldots q_{i+p-1}}_{=Q_{i,p}} \cdot \underbrace{q_{i+p} q_{i+p+1} \ldots q_{i+k-1}}_{=Q_{i+p,k-p} = Q_{i,k-p}} = Q_{i,p} \cdot Q_{i,k-p}. \tag{13}$$

The $i = 0$ case is $Q_{0,k} = Q_{0,p} \cdot Q_{0,k-p}$; dividing this by (13) gives

$$\frac{Q_{0,k}}{Q_{i,k}} = \frac{Q_{0,p}}{Q_{i,p}} \cdot \frac{Q_{0,k-p}}{Q_{i,k-p}} = \frac{Q_{0,k-p}}{Q_{i,k-p}} \qquad \text{for } k \geqslant p, \tag{14}$$

where we used $Q_{i,p} = \prod_{j=i}^{i+p-1} q_{j \bmod p} = Q_{0,p}$ (same factors in a different order). Iterating (14) gives $Q_{0,k}/Q_{i,k} = Q_{0,k'}/Q_{i,k'}$. ☐

Thus, (12) is equivalent to $m_i = (Q_{0,k}/Q_{i,k})m_0 = (Q_{0,k'}/Q_{i,k'})m_0$, with $k' = k \bmod p$. Then

$$\vec{m} = m_0 \cdot \left(1, \frac{\theta_0}{\theta_1}, \frac{\theta_0}{\theta_2}, \dots, \frac{\theta_0}{\theta_{p-1}}\right) = m_0 \cdot \left(1, \frac{Q_{0,k'}}{Q_{1,k'}}, \frac{Q_{0,k'}}{Q_{2,k'}}, \dots, \frac{Q_{0,k'}}{Q_{p-1,k'}}\right). \tag{15}$$

Alternatively, expand $m_{i-1}\theta_{i-1} = m_i\theta_i$ as

$$m_{i-1}q_{i-1}q_iq_{i+1}\dots q_{i+k-2} = m_iq_iq_{i+1}\dots q_{i+k-1}\ , \tag{16}$$

and cancel common factors $q_i \dots q_{i+k-2}$, to obtain

$$m_{i-1}q_{i-1} = m_iq_{i+k-1} \qquad \text{for all } i \in \mathbb{P}. \tag{17}$$

This is equivalent to (10): parameters are compatible iff they are positive integers satisfying (17).

Eq. (17) gives $m_i = (q_{i-1}/q_{i+k-1})m_{i-1}$. Now, let $r \geqslant 0$ be an integer (in $\mathbb{Z}$, not $\mathbb{P}$). Iterating this r times starting with m_r gives

$$m_r = (Q_{0,r}/Q_{k,r})m_0 \qquad \text{for all integers } r \geqslant 0 \text{ (note } r \in \mathbb{Z} \text{ rather than } r \in \mathbb{P}). \tag{18}$$

To use (18) to compute m_i for $i \in \mathbb{P}$, pick any $r \geqslant 0$ in $\mathbb{Z}$ with $r \equiv i \pmod p$, and evaluate (18) at r. Consider two such choices, r_1 and r_2; by Lemma 3, $Q_{0,r_1}/Q_{k,r_1} = Q_{0,r_2}/Q_{k,r_2}$ so both give the same result on the right side of (18). Then

$$\vec{m} = m_0 \cdot \left(1, \frac{Q_{0,1}}{Q_{k,1}}, \frac{Q_{0,2}}{Q_{k,2}}, \cdots, \frac{Q_{0,p-1}}{Q_{k,p-1}}\right). \tag{19}$$

On plugging in positive integers for k, p, q_i's, and m_0, the components of this vector are rational, but may not be integers. To find all compatible parameters:

(i) Pick any positive integers k, p, and $q_0, \dots, q_{p-1}$, and plug them into (19).

(ii) For $i = 0, \dots, p-1$, reduce fractions $Q_{0,i}/Q_{k,i}$ to lowest terms, a_i/b_i.

(iii) Set m_0 to any positive integer multiple of the least common multiple of the denominators $b_0, \dots, b_{p-1}$ after such reduction.

Alternatively, we could use (15) instead of (19), and θ_0/θ_i or $Q_{0,k'}/Q_{i,k'}$ instead of $Q_{0,i}/Q_{k,i}$ (all three formulas are equal, so the results are the same). We give an example, using (19).

Example 4. Let $p = 3$ and $\vec{q} = (2, 4, 6)$. There are $p = 3$ cases, depending on $k \bmod 3$:

	Step (i)	(ii)	(iii)	
if $k \equiv 0 \pmod 3$:	$\vec{m} = m_0 \cdot \left(1, \frac{2}{2}, \frac{2\cdot 4}{2\cdot 4}\right)$	$= m_0 \cdot (1,1,1)$	$= m_0 \cdot (1,1,1)$	with $m_0 \in \mathbb{Z}^+$;
if $k \equiv 1$:	$\vec{m} = m_0 \cdot \left(1, \frac{2}{4}, \frac{2\cdot 4}{4\cdot 6}\right)$	$= m_0 \cdot \left(1, \frac{1}{2}, \frac{1}{3}\right)$	$= \frac{m_0}{6} \cdot (6,3,2)$	with $6 \mid m_0$;
if $k \equiv 2$:	$\vec{m} = m_0 \cdot \left(1, \frac{2}{6}, \frac{2\cdot 4}{6\cdot 2}\right)$	$= m_0 \cdot \left(1, \frac{1}{3}, \frac{2}{3}\right)$	$= \frac{m_0}{3} \cdot (3,1,2)$	with $3 \mid m_0$.

For any $q_0 \in \mathbb{Z}^+$, taking $\vec{q} = q_0 \cdot (1, 2, 3)$ will also give these same solutions for $\vec{m}$.

The $k \equiv 0 \pmod 3$ example above generalizes:

Lemma 5. *Let $p, k, q_0, \ldots, q_{p-1} \in \mathbb{Z}^+$ and $k \equiv 0 \pmod p$. Then $(\vec{m}, \vec{q}, k)$ are compatible parameters iff $m_0 = \cdots = m_{p-1} \in \mathbb{Z}^+$ iff $\vec{m} = m_0 \cdot (1, \ldots, 1)$ for some $m_0 \in \mathbb{Z}^+$.*

Proof. Since $k \equiv 0 \pmod p$, we have $Q_{k,r} = Q_{0,r}$ for all integers $r \geqslant 0$. So by (18), $m_r = m_0 Q_{0,r}/Q_{k,r} = m_0$ for all $r \geqslant 0$. Thus, $\vec{m} = (m_0, \ldots, m_0) = m_0 \cdot (1, \ldots, 1)$. □

When $(\vec{m}, \vec{q}, k)$ are compatible parameters, certain multiples of $\vec{m}$ and $\vec{q}$ also give compatible parameters; this was illustrated in Example 4, and generalizes as follows.

For $h \in \mathbb{Z}^+$, let $h\,\vec{m} = (h \cdot m_0, \ldots, h \cdot m_{p-1})$ and $\vec{m}/h = (m_0/h, \ldots, m_{p-1}/h)$.

Lemma 6. *Let $(\vec{m}, \vec{q}, k)$ be compatible parameters. Then the following are also compatible:*

(a) $(h\,\vec{m}, \vec{q}, k)$ *for* $h \in \mathbb{Z}^+$;
(b) $(\vec{m}, h\,\vec{q}, k)$ *for* $h \in \mathbb{Z}^+$;
(c) $(\vec{m}/h, \vec{q}, k)$ *for* $h \mid \gcd(m_0, \ldots, m_{p-1})$;
(d) $(\vec{m}, \vec{q}/h, k)$ *for* $h \mid \gcd(q_0, \ldots, q_{p-1})$;
(e) $(\rho(\vec{m}), \rho(\vec{q}), k)$.

Proof. Given positive integers p, k, $m_0, \ldots, m_{p-1}$, and $q_0, \ldots, q_{p-1}$ satisfying (17):

(a) Multiplying both sides of (17) by h gives $(h\,m_{i-1}) \cdot q_{i-1} = (h\,m_i) \cdot q_{i+k-1}$ for all $i \in \mathbb{P}$. Thus, $(h\,\vec{m}, \vec{q}, k)$ also satisfies (17) in positive integers.

(c) Dividing both sides of (17) by h gives $(m_{i-1}/h)q_{i-1} = (m_i/h)q_{i+k-1}$ for all $i \in \mathbb{P}$. Then $(\vec{m}/h, \vec{q}, k)$ also satisfies (17) in positive integers.

(b,d) Similar proofs to (a) and (c), but rescaling q's instead of m's.

(e) Applying (17) to $(\rho(\vec{m}), \rho(\vec{q}), k)$ shifts i to $i-1$, yielding the same p equations as (17) but indexed differently. □

Lemma 7. *If $(\vec{m}, \vec{q}, k)$ are compatible parameters and if $q_i \geqslant 2$ for at least one i, then $\ell > k$. In fact, $\ell \geqslant (e \cdot \ln(2)) \cdot k$, where $e \cdot \ln(2) \approx 1.884169$.*

Proof. Choose any i in $0 \leqslant i \leqslant p-1$ with $q_i \geqslant 2$. By (11), $\ell = p \cdot m_i \cdot q_i \cdots q_{i+k-1}$. All factors in this product are positive integers, and hence at least 1. Since the sequence of q's repeats every p factors, the factor q_i is repeated $\lceil k/p \rceil$ times. So $\ell \geqslant p \cdot 1 \cdot 2^{\lceil k/p \rceil} \geqslant p \cdot 2^{k/p}$.

Consider $f(x) = 2^x/x$ for $x \in \mathbb{R}^+$. This has an absolute minimum at $x = 1/\ln(2)$, with value $f(1/\ln(2)) = e \cdot \ln(2) \approx 1.884169 > 1$. So $f(x) > 1$ for all $x > 0$.

Then $f(k/p) = (p/k)2^{k/p} > 1$, so $p \cdot 2^{k/p} > k$. Combining with $\ell \geqslant p \cdot 2^{k/p}$ gives $\ell > k$.

For the stronger result, $(p/k)2^{k/p} \geqslant e \cdot \ln(2)$, so $\ell \geqslant p \cdot 2^{k/p} \geqslant (e \cdot \ln(2)) \cdot k$. □

3 Phased Multi de Bruijn Graph

3.1 Spelling strings by walks in a graph

In this section, we use the method of de Bruijn graphs to form a directed multigraph $G = G(\vec{m}, \vec{q}, k)$. We require p, k, $m_0, \ldots, m_{p-1}$, and $q_0, \ldots, q_{p-1}$ to be positive integers (but not necessarily compatible parameters).

The set V of vertices is defined in (6); it's the set of phased $(k-1)$-mers in all phases.

For each k-mer $y = c_0c_1 \ldots c_{k-1} \in K_i$, add m_i directed edges labelled y on the same two vertices:

$$\underbrace{c_0c_1 \ldots c_{k-2}}_{\text{phase } i} \xrightarrow{y} \underbrace{c_1c_2 \ldots c_{k-1}}_{\text{phase } i+1} . \tag{20}$$

Every walk in G has the following form:

$$b_0b_1 \cdots b_{k-2} \;\rightarrow\; b_1b_2 \cdots b_{k-1} \;\rightarrow\; b_2b_3 \cdots b_k \;\rightarrow\; \cdots \;\rightarrow\; b_rb_{r+1} \cdots b_{r+k-2} . \tag{21}$$

Here, $r \geqslant 0$ is the length of the walk (in edges); $i_0 \in \mathbb{P}$ is the phase of the starting vertex; and $b_j \in \Omega_{i_0+j}$ for integers $j = 0, \ldots, r+k-2$.

The *linear string* of this walk is $b_0b_1 \ldots b_{r+k-2}$, of length $r+k-1$ characters; this can be formed using vertex labels (if $k \geqslant 2$) or edge labels (if $k \geqslant 1$). It is a phased string, whose phase is the same as the first vertex (or edge). If the walk is closed (first and last vertex are the same), then its *linearized string* is $b_0 \cdots b_{r-1}$ and its *cyclic string* is $(b_0b_1 \cdots b_{r-1})$, each of length r. Starting at another location in the walk represents the same cyclic sequence by a cyclic shift.

For a closed walk based at vertex $v = b_0b_1 \ldots b_{k-2}$, the relation between its linearized string $s = b_0b_1 \ldots b_{r-1}$ and its linear string $s' = b_0b_1 \ldots b_{r-k+2}$ is $s' = sv$ (concatenation of s and v): the final vertex of a closed walk equals the initial vertex, v, and both represent the same $(k-1)$-mer.

Let $\mathcal{L}(\vec{m}, \vec{q}, k)$ be the set of phased linear sequences in which every element of K_i occurs exactly m_i times, for each $i \in \mathbb{P}$. Any $s \in \mathcal{L}(\vec{m}, \vec{q}, k)$ can be represented by an Eulerian trail in G (every edge used exactly once) of form (21); the m_i occurrences in s of each k-mer $y \in K_i$ are represented by assigning each to one of the m_i edges of G labelled by y. Conversely, the linear string of any Eulerian trail in G gives a sequence in $\mathcal{L}(\vec{m}, \vec{q}, k)$. Similarly, linearized strings $s \in \mathcal{LC}(\vec{m}, \vec{q}, k)$ and cyclic strings $(s) \in \mathcal{C}(\vec{m}, \vec{q}, k)$ are represented by Eulerian cycles in G (closed walks using every edge exactly once).

3.2 Adjacency matrix

For any string $s = s_0s_1 \ldots s_{r-1}$ of length $r \geqslant 1$, and any integer h in $0 \leqslant h \leqslant r$, define the *prefix* and *suffix* of s as the first or last h characters of s, respectively:

$$\textsc{Pre}_h(s) = s_0s_1 \ldots s_{h-1} \qquad \textsc{Suf}_h(s) = s_{r-h}s_{r-h+1} \ldots s_{r-1} \tag{22}$$

$$\textsc{Pre}(s) = s_0s_1 \ldots s_{r-2} \qquad \textsc{Suf}(s) = s_1s_2 \ldots s_{r-1} \tag{23}$$

(if the subscript h is omitted, use $h = r-1$ to remove just one character).

Next, we develop the adjacency matrix of G. First, let $k \geqslant 2$. For each k-mer $y = y_0 \ldots y_{k-1} \in K_i$, the edges (20) labelled by y have the form $v \to w$, where $v = \text{Pre}(y) = y_0 \ldots y_{k-2}$ has phase i and $w = \text{Suf}(y) = y_1 \ldots y_{k-1}$ has phase $i+1$. These also have $\text{Suf}(v) = \text{Pre}(w) = y_1 \ldots y_{k-2}$, and edge multiplicity m_i. For $i, j \in \mathbb{P}$ and each pair of vertices $v \in V_i$ and $w \in V_j$, set

$$\text{For } k \geqslant 2: \quad A_{v,w} = \begin{cases} m_i & \text{if } j \equiv i+1 \pmod{p} \\ & \text{and } \text{Suf}(v) = \text{Pre}(w); \\ 0 & \text{otherwise.} \end{cases} \tag{24}$$

Since $\text{Suf}(v) = \text{Pre}(w)$ implies $\text{Phase}(w) = \text{Phase}(v) + 1$, it's not strictly necessary to state $j \equiv i+1 \pmod{p}$ in (24).

Next, $k = 1$ is a special case. A phase i 1-mer has the form $y = y_0 \in \Omega_i$, which gives vertices $v = \text{Pre}(y) = \emptyset_i$ and $w = \text{Suf}(y) = \emptyset_{i+1}$. Each of the q_i choices of $y_0 \in \Omega_i$ yields m_i edges $\emptyset_i \xrightarrow{y_0} \emptyset_{i+1}$; thus there are $m_i q_i$ edges from $\emptyset_i$ to $\emptyset_{i+1}$. Then

$$\text{For } k = 1: \quad A_{v,w} = \begin{cases} m_i\, q_i & \text{if } j \equiv i+1 \pmod{p} \\ 0 & \text{otherwise.} \end{cases} \tag{25}$$

Classical de Bruijn sequences are the case $p = 1$, $m_0 = 1$, $q_0 \geqslant 2$, $k \geqslant 1$. In different notation, Dawson and Good [4] made a stochastic transition matrix T and showed that T^{k-1} has constant columns, which leads to the eigenvalues of T and of the Laplacian matrix (to be defined in Sec. 4.1).

In [18], for the case $p = 1$, $m_0 \geqslant 1$, $q_0 \geqslant 1$, $k \geqslant 2$, we showed that all entries of A^{k-1} equal ${m_0}^{k-1}$, which also leads to the eigenvalues of A and of the Laplacian matrix.

We realized the method in [4, 18] to evaluate A^{k-1} generalizes as follows (Lemma 8), which lead us to study phased multi de Bruijn sequences; however, A^{k-1} does not have constant columns. We'll use a different method in Sec. 7 to evaluate the eigenvalues of the Laplacian matrix.

Lemma 8. *Let $v \in V_i$, $w \in V_j$, and $r \geqslant 0$. The number of directed walks in G of length r from v to w is*

$$(A^r)_{v,w} = \begin{cases} m_i m_{i+1} \ldots m_{i+r-1} Q_{i+k-1,r-(k-1)} & \textit{if } r \geqslant k-1 \\ & \textit{and } r \equiv j - i \pmod{p}; \\ m_i m_{i+1} \ldots m_{i+r-1} & \textit{if } 0 \leqslant r \leqslant k-1 \\ & \textit{and } r \equiv j-i \pmod{p} \\ & \textit{and } \text{Suf}_{k-1-r}(v) = \text{Pre}_{k-1-r}(w); \\ 0 & \textit{otherwise.} \end{cases} \tag{26}$$

Proof. The number of directed walks in G of length r from v to w is $(A^r)_{v,w}$. Each directed edge in the graph advances by one phase, so $(A^r)_{v,w} = 0$ when $r \not\equiv j - i \pmod{p}$. For the rest of this proof, assume $r \equiv j - i \pmod{p}$.

Set $v = b_0 b_1 \cdots b_{k-2} \in V_i$ and $w = c_0 c_1 \cdots c_{k-2} \in V_j$.

Case $r \geqslant k-1$*:* Select $x_h \in \Omega_{i+k-1+h}$ for $h = 0, \ldots, r-k$. There are $Q_{i+k-1,r-k+1} > 0$ ways to choose the x's. Then set

$$s = b_0 b_1 \cdots b_{k-2}\, x_0 x_1 \ldots x_{r-k}\, c_0 c_1 \cdots c_{k-2} \in \mathcal{S}_{i,r+k-1}\ . \tag{27}$$

The sequence of $(k-1)$-mers starting at successive positions of s gives vertices along a directed walk from v to w. There are parallel edges for each consecutive pair of vertices, with edge multiplicities $m_i, m_{i+1}, \ldots, m_{i+r-1}$ indicated on each arrow:

$$b_0 b_1 \cdots b_{k-2} \xrightarrow{m_i} b_1 b_2 \cdots b_{k-2} x_0 \xrightarrow{m_{i+1}} \cdots \xrightarrow{m_{i+r-2}} x_{r-k} c_0 c_1 \cdots c_{k-3} \xrightarrow{m_{i+r-1}} c_0 c_1 \cdots c_{k-2}.$$

Thus, $m_i m_{i+1} \cdots m_{i+r-1}$ directed walks spell the sequence s. Combining with the number of choices of x's gives $m_i m_{i+1} \ldots m_{i+r-1} \cdot Q_{i+k-1,r-k+1}$ directed walks from v to w.

Note: We may consider $r = k-1$ with this case or the next case. Here, when $r = k-1$, there are no x's and we have $s = b_0 b_1 \cdots b_{k-2} c_0 c_1 \cdots c_{k-2}$ and $Q_{i+k-1,r-k+1} = Q_{i+k-1,0} = 1$.

Case $0 \leqslant r \leqslant k-1$*:* Each directed edge shifts out a character from the left end of the vertex and adds a new character to the right end. After r steps, the last $k-1-r \geqslant 0$ characters of the initial vertex must match the first $k-1-r$ characters of the final vertex; that is, $b_r b_{r+1} \ldots b_{k-2} = c_0 c_1 \ldots c_{k-2-r}$, or more succinctly, $\textsc{Suf}_{k-r-1}(v) = \textsc{Pre}_{k-1-r}(w)$.

If $\textsc{Suf}_{k-r-1}(v) \neq \textsc{Pre}_{k-1-r}(w)$, there are no walks of length r from v to w: $(A^r)_{v,w} = 0$.

If $\textsc{Suf}_{k-r-1}(v) = \textsc{Pre}_{k-1-r}(w)$, then merge v and w by overlapping the positions of these $k-1-r$ characters, to form a string s of length $2(k-1)-(k-1-r) = r+k-1$:

$$s = b_0 b_1 \ldots b_{r-1}\ \underbrace{b_r b_{r+1} \ldots b_{k-2}}_{= c_0 c_1 \ldots c_{k-r-2}}\ c_{k-r-1} c_{k-r} \ldots c_{k-2} \in \mathcal{S}_{i,r+k-1}\ .$$

Any r step walk from v to w must spell the string s and must have this form:

$$b_0 b_1 \cdots b_{k-2} \xrightarrow{m_i} b_1 b_2 \cdots b_{k-2} c_{k-r-1} \xrightarrow{m_{i+1}} \cdots \xrightarrow{m_{i+r-2}} b_{r-1} c_0 c_1 \cdots c_{k-3} \xrightarrow{m_{i+r-1}} c_0 c_1 \cdots c_{k-2}.$$

Choosing among parallel edges for each step again leads to $m_i m_{i+1} \ldots m_{i+r-1}$ directed walks. But unlike the previous case, s is fully determined from v, w, and r (there are no x_h's to choose), so this is the total number of walks, giving the middle case of (26).

Note: If $r = k-1$, then $k-1-r = 0$ and $s = b_0 b_1 \cdots b_{k-2} c_0 c_1 \cdots c_{k-2}$. Also, $\textsc{Suf}_{k-r-1}(v) = \textsc{Suf}_0(v) = \emptyset_{i+k-1}$ and $\textsc{Pre}_{k-r-1}(w) = \textsc{Pre}_0(w) = \emptyset_j$. These phased null strings are equal iff their phases agree: $i+k-1 \equiv j \pmod{p}$, which holds iff $k-1 \equiv j-i \pmod{p}$ iff $r \equiv j-i \pmod{p}$. □

Corollary 9. *Let $k, p, m_0, \ldots, m_{p-1}, q_0, \ldots, q_{p-1} \in \mathbb{Z}^+$. Then $G = G(\vec{m}, \vec{q}, k)$ is strongly connected.*

Proof. Let v and w be any two vertices, and i and j be their phases. Take any integer $r \geqslant k-1$ such that $r \equiv j-i \pmod{p}$. The number of directed walks of length r from v to w is $(A^r)_{v,w}$; this is positive by the first case in (26), so there is at least one directed walk from v to w. This holds for any vertices v and w, so G is strongly connected. □

3.3 Degrees

This section discusses indegrees and outdegrees of vertices.

Lemma 10. *Consider the graph* $G = G(\vec{m}, \vec{q}, k)$*. For all parameters* $(\vec{m}, \vec{q}, k)$ *(whether or not compatible), every vertex* $v \in V_i$ *has* $\textsc{Indeg}(v) = m_{i-1}q_{i-1}$ *and* $\textsc{Outdeg}(v) = m_i q_{i+k-1}$*. Graph G is balanced (every vertex has equal indegree and outdegree) iff* $(\vec{m}, \vec{q}, k)$ *are compatible parameters.*

Proof. First consider $k \geqslant 3$. Let $v = c_0 \ldots c_{k-2} \in V_i$. For each $x \in \Omega_{i-1}$, there are m_{i-1} parallel incoming edges of the following form, so $\textsc{Indeg}(v) = m_{i-1}q_{i-1}$:

$$xc_0 \ldots c_{k-3} \xrightarrow{xc_0 \ldots c_{k-2}} v.$$

Similarly, for each $y \in \Omega_{i+k-1}$, there are m_i parallel outgoing edges of the following form, so $\textsc{Outdeg}(v) = m_i q_{i+k-1}$:

$$v \xrightarrow{c_0 \ldots c_{k-2} y} c_1 \ldots c_{k-2}\, y.$$

This argument also works for $k = 2$ by using $c_0 \ldots c_{k-3} = \emptyset_i$ and $c_1 \ldots c_{k-2} = \emptyset_{i+1}$.

For $k = 1$, the vertices are $V = \{\emptyset_0, \ldots, \emptyset_{p-1}\}$. The incoming edges to $\emptyset_i$ are $\emptyset_{i-1} \xrightarrow{c} \emptyset_i$ for each $c \in \Omega_{i-1}$, with multiplicity m_{i-1}, so $\textsc{Indeg}(\emptyset_i) = m_{i-1}q_{i-1}$. The outgoing edges from $\emptyset_i$ are $\emptyset_i \xrightarrow{c} \emptyset_{i+1}$ for each $c \in \Omega_i$, with multiplicity m_i, so $\textsc{Outdeg}(\emptyset_i) = m_i q_i = m_i q_{i+k-1}$ (since $k = 1$).

Finally, the graph is balanced iff every vertex has equal in and outdegree, iff $m_{i-1}q_{i-1} = m_i q_{i+k-1}$ for all $i \in \mathbb{P}$. By (17), this is equivalent to $(\vec{m}, \vec{q}, k)$ being compatible. $\square$

Define

$$d_i = m_{i-1}q_{i-1} \qquad \text{for } i \in \mathbb{P}. \tag{28}$$

For all parameters $(\vec{m}, \vec{q}, k)$, this is the indegree of every phase i vertex in $G(\vec{m}, \vec{q}, k)$. For compatible parameters, it's also the outdegree, giving this extension of (17):

$$d_i = m_{i-1}q_{i-1} = m_i q_{i+k-1} \qquad \text{for } i \in \mathbb{P}. \tag{29}$$

Corollary 11. $\mathcal{C}(\vec{m}, \vec{q}, k)$ *is nonempty iff* $(\vec{m}, \vec{q}, k)$ *are compatible parameters.*

Proof. As explained in Sec. 3.1, each element of $\mathcal{C}(\vec{m}, \vec{q}, k)$ is represented by one or more Eulerian cycles in $G(\vec{m}, \vec{q}, k)$, and each Eulerian cycle spells out an element of $\mathcal{C}(\vec{m}, \vec{q}, k)$. However, G only has Eulerian cycles iff it is strongly connected and balanced. The graph is strongly connected by Cor. 9 for all positive integer parameters p, k, $m_0, \ldots, m_{p-1}$, and $q_0, \ldots, q_{p-1}$. By Lemma 10, it is balanced iff $(\vec{m}, \vec{q}, k)$ are compatible. Thus, $\mathcal{C}(\vec{m}, \vec{q}, k)$ is nonempty iff $(\vec{m}, \vec{q}, k)$ are compatible. $\square$

In view of Cor. 11, the remaining sections of this paper assume that $(\vec{m}, \vec{q}, k)$ are compatible parameters unless otherwise stated.

4 Counting cyclic phased multi de Bruijn sequences

In this section, we will count Eulerian cycles in G, and account for multiple Eulerian cycles representing each element of $\mathcal{C}(\vec{m}, \vec{q}, k)$, where $(\vec{m}, \vec{q}, k)$ are compatible parameters.

4.1 Laplacian

The *degree matrix* of G is a diagonal matrix D indexed by $V \times V$. For $i, j \in \mathbb{P}$ and each pair of vertices $v \in V_i$ and $w \in V_j$, set

$$D_{v,w} = \begin{cases} d_i & \text{if } v = w \in V_i; \\ 0 & \text{otherwise.} \end{cases} \tag{30}$$

The *Laplacian matrix* of G is $L = D - A$. For $k \geqslant 2$, compute it using (30) and (24) (the case $k = 1$ is deferred to Theorem 27): When $p \geqslant 2$, the diagonal of A is 0 since $A_{v,w} \neq 0$ requires v and w to have consecutive phases; thus, L matches D on the diagonal and $-A$ off-diagonal. But for $p = 1$, there are nonzero diagonal entries $A_{v,v} = m_i$ for $v = a^{k-1}$ (with $a \in \Omega_i$ and $i = 0$). Thus:

$$\text{For } p \geqslant 2 \text{ and } k \geqslant 2: \quad L_{v,w} = \begin{cases} d_i & \text{if } v = w \in V_i; \\ -m_i & \text{if } \textsc{Suf}(v) = \textsc{Pre}(w); \\ 0 & \text{otherwise.} \end{cases} \tag{31}$$

$$\text{For } p = 1 \text{ and } k \geqslant 2: \quad L_{v,w} = \begin{cases} d_i - m_i & \text{if } v = w = a^{k-1} \text{ for any } a \in \Omega_0; \\ d_i & \text{all other cases with } v = w; \\ -m_i & \text{if } \textsc{Suf}(v) = \textsc{Pre}(w) \text{ but } v \neq w; \\ 0 & \text{otherwise.} \end{cases} \tag{32}$$

For $k = 1$: See Theorem 27.

The number of Eulerian cycles in G can be computed using the eigenvalues of L, via the Matrix-Tree and BEST Theorems. We will use the characteristic polynomial of L to analyze the eigenvalues:

Theorem 12. *For compatible parameters $(\vec{m}, \vec{q}, k)$, the characteristic polynomial of L is*

$$\det(xI - L) = \left(\prod_{i=0}^{p-1} (x - d_i)^{n_i - 1} \right) \cdot \left(\prod_{i=0}^{p-1} (x - d_i) - \prod_{i=0}^{p-1} (-d_i) \right) . \tag{33}$$

The proof of Theorem 12 is rather long; we postpone it to Sec. 7. See Appendix A.2 for a worked-out example (with $p = 4$ phases) for Theorems 12 and 13.

Theorem 13. *For compatible parameters $(\vec{m}, \vec{q}, k)$, the eigenvalues of L are described as follows, and each distinct eigenvalue is in just one of (a), (b), or (c):*

(a) *For each* $i \in \mathbb{P}$*, phase* i *contributes eigenvalue* d_i *with multiplicity* $n_i - 1$*. If different phases have the same value of* d_i*, add their multiplicities together for the total multiplicity of* d_i*. This accounts for* $\sum_{i=0}^{p-1}(n_i - 1) = n - p$ *of the eigenvalues.*

(b) *There is one eigenvalue of* 0*.*

(c) *The remaining* $p - 1$ *eigenvalues have product*

$$d_0 d_1 \cdots d_{p-1} \cdot \left(\frac{1}{d_0} + \cdots + \frac{1}{d_{p-1}}\right). \tag{34}$$

Proof. Let $f(x) = \det(xI - L)$, given by (33). Factors $\prod_{i=0}^{p-1}(x - d_i)^{n_i - 1}$ contribute the $n - p$ eigenvalues listed in (a). The other factor in (33) is a polynomial in x of degree p:

$$g(x) = \prod_{i=0}^{p-1}(x - d_i) - \prod_{i=0}^{p-1}(-d_i).$$

Since $g(0) = 0$, at least one eigenvalue is 0. The product of the remaining $p-1$ eigenvalues of L is $(-1)^{p-1}$ times the constant term of $g(x)/x$; that is, $(-1)^{p-1}g'(0)$:

$$g'(x) = \left(\prod_{i=0}^{p-1}(x - d_i)\right) \cdot \sum_{i=0}^{p-1} \frac{1}{x - d_i}$$
$$g'(0) = \left(\prod_{i=0}^{p-1}(-d_i)\right) \cdot \sum_{i=0}^{p-1} \frac{1}{-d_i} = (-1)^{p+1} \left(\prod_{i=0}^{p-1} d_i\right) \cdot \sum_{i=0}^{p-1} \frac{1}{d_i}$$

Then $(-1)^{p-1}g'(0)$ is given by (34), proving (c). It is nonzero (as all d_i are positive), so there are no additional eigenvalues of 0, thus establishing (b).

Parts (a) and (c) give nonzero eigenvalues, so they're distinct from (b). If d_j is an eigenvalue listed in (a), then $g(d_j) = 0 - \prod_{i=0}(-d_i) \neq 0$, so d_j is not contributed as an eigenvalue in (c). Thus, each distinct eigenvalue is only in one of (a), (b), or (c). □

Lemma 14. *For compatible parameters* $(\vec{m}, \vec{q}, k)$*, the number of phase* i *vertices is:*

$$n_i = \frac{\ell}{p\, d_i} \tag{35}$$

and the total number of vertices is

$$n = \frac{\ell}{p}\left(\frac{1}{d_0} + \cdots + \frac{1}{d_{p-1}}\right). \tag{36}$$

Proof. Compute n_i using (4)–(5), (11), and (29):

$$n_i = |V_i| = Q_{i,k-1} = q_i q_{i+1} \cdots q_{i+k-2} = \frac{\ell}{p\, m_i\, q_{i+k-1}} = \frac{\ell}{p\, d_i}.$$

Then $n = |V| = n_0 + \cdots + n_{p-1}$ gives (36). □

Corollary 15. *For compatible parameters $(\vec{m}, \vec{q}, k)$, the product of the $n-1$ nonzero eigenvalues of L is*

$$\frac{np}{\ell} \cdot {d_0}^{n_0} \cdots {d_{p-1}}^{n_{p-1}}. \tag{37}$$

Proof. Multiplying the eigenvalues (with multiplicities) from Lemma 13(a,c) gives

$$\left(\prod_{i=0}^{p-1} {d_i}^{(n_i-1)+1}\right) \cdot \left(\sum_{i=0}^{p-1} \frac{1}{d_i}\right) = {d_0}^{n_0} \cdots {d_{p-1}}^{n_{p-1}} \cdot \underbrace{\left(\frac{1}{d_0} + \cdots + \frac{1}{d_{p-1}}\right)}_{=np/\ell \text{ by } (36)}. \qquad \square$$

4.2 Number of Eulerian cycles

Let $e = (v, w)$ be any edge in G. This edge represents some k-mer $y \in K_i$, so v is in phase i and w is in phase $i+1$.

First, we count how many spanning trees of G have root v (all tree edges directed towards v). We use Tutte's Matrix-Tree Theorem ([19, Thm. 3.6]; also see [20, Thm. 7] and [17, Thm. 5.6.4]); specifically, a version for balanced directed graphs ([17, Cor. 5.6.6]):

$$\begin{aligned} \#\text{ spanning trees rooted at } v &= (1/n) \cdot (\text{product of all eigenvalues of } L \text{ except one } 0) \\ &= (p/\ell) \cdot {d_0}^{n_0} \cdots {d_{p-1}}^{n_{p-1}}, \end{aligned} \tag{38}$$

where we used (37). Next, we compute the number of Eulerian cycles in G starting with edge e. By the BEST Theorem [20, Theorem 5b] (also see [17, pp. 56, 68] and [8]), this is

$$\begin{aligned} &(\#\text{ spanning trees rooted at } v) \cdot \prod_{x \in V} (\textsc{Outdeg}(x) - 1)! \\ &= \frac{p}{\ell} \cdot {d_0}^{n_0} \cdots {d_{p-1}}^{n_{p-1}} \cdot \prod_{j=0}^{p-1} (d_j - 1)!^{n_j} = \frac{p}{\ell} \cdot \prod_{j=0}^{p-1} (d_j!)^{n_j}. \end{aligned} \tag{39}$$

Each Eulerian cycle spells out a linearization of a sequence in $\mathcal{C}(\vec{m}, \vec{q}, k)$, starting with k-mer y. However, due to edge multiplicities, there may be multiple Eulerian cycles yielding each linearized sequence. Let C be an Eulerian cycle that starts on edge e and spells out a linearization s:

- The initial edge e is given. The $m_i - 1$ edges parallel to e also represent k-mer y, and may be permuted in C in $(m_i - 1)!$ ways.
- Let $j \in \mathbb{P}$ and $x \in K_j$, with $x \neq y$. Then x occurs m_j times in s, and the edges representing x may be permuted in C in $m_j!$ ways. There are $\theta_i - 1$ choices of x if $j = i$, and θ_j if $j \neq i$.
- Thus, the number of Eulerian cycles that start on edge e and give linearization s is

$$(m_i - 1)! \cdot (m_i!)^{\theta_i - 1} \cdot \prod_{j \neq i} (m_j!)^{\theta_j} = \frac{1}{m_i} \prod_{j=0}^{p-1} (m_j!)^{\theta_j}. \tag{40}$$

Recall that $\mathcal{LC}_y(\vec{m},\vec{q},k)$ is the set of linearizations of sequences in $\mathcal{C}(\vec{m},\vec{q},k)$ that start with k-mer y. Compute its size by dividing (39) by (40):

$$|\mathcal{LC}_y(\vec{m},\vec{q},k)| = \frac{p\,m_i}{\ell}\cdot\frac{\prod_{j=0}^{p-1}(d_j!)^{n_j}}{\prod_{j=0}^{p-1}(m_j!)^{\theta_j}} = \frac{p\,m_i}{\ell}\cdot\frac{\prod_{j=0}^{p-1}(d_{j+1}!)^{n_{j+1}}}{\prod_{j=0}^{p-1}(m_j!)^{\theta_j}}. \tag{41}$$

Note that this depends on y via $i = \textsc{Phase}(y)$.

Next, use $d_{j+1} = m_j q_j$ (by (28)); $\theta_j = q_j n_{j+1}$ (by (5)); and $p\,m_i/\ell = 1/\theta_i$ (by (11)):

$$|\mathcal{LC}_y(\vec{m},\vec{q},k)| = \frac{1}{\theta_i}\cdot\prod_{j=0}^{p-1}\frac{\Big((m_jq_j)!\Big)^{n_{j+1}}}{(m_j!)^{q_jn_{j+1}}} = \frac{1}{\theta_i}\cdot\prod_{j=0}^{p-1}\left(\frac{(m_jq_j)!}{(m_j!)^{q_j}}\right)^{n_{j+1}} = \frac{W(\vec{m},\vec{q},k)}{\theta_i}\,, \tag{42}$$

where we define the following (note that (5) gives n_{j+1} in terms of $\vec{m},\vec{q},k$):

$$W(\vec{m},\vec{q},k) = \prod_{j=0}^{p-1}\left(\frac{(m_jq_j)!}{(m_j!)^{q_j}}\right)^{n_{j+1}}. \tag{43}$$

Denote the set of linearized sequences in phase i by $\mathcal{LC}_{[[i]]}(\vec{m},\vec{q},k)$. The number of such sequences is Eq. (42) multiplied by θ_i choices of $y \in K_i$:

$$|\mathcal{LC}_{[[i]]}(\vec{m},\vec{q},k)| = W(\vec{m},\vec{q},k). \tag{44}$$

This is equal for all i since each $(s) \in \mathcal{C}(\vec{m},\vec{q},k)$ has an equal number of distinct linearizations per phase (namely $\ell/(pd)$, where (s) has rotational order d). Multiply (44) by p to count linearized sequences in all phases:

$$|\mathcal{LC}(\vec{m},\vec{q},k)| = p\cdot W(\vec{m},\vec{q},k). \tag{45}$$

Parameters for classical and multi de Bruijn sequences are given in Table 2. Classical de Bruijn sequences have a unique linearization starting with any $y \in K$ (corresponding to initial edge e in the BEST Theorem). The number of classical de Bruijn sequences is given by plugging Table 2(a) into (42):

$$|\mathcal{C}((1),(q),k)| = |\mathcal{LC}_y((1),(q),k)| = \frac{(q!)^{q^{k-1}}}{q^k}. \tag{46}$$

This matches the result (with different notation) by van Aardenne-Ehrenfest and de Bruijn [20, p. 203] for $q,k \geqslant 1$, and by Sainte-Marie [15] and de Bruijn [5] for $q = 2$.

Plugging the values from Table 2(b) into (42) gives the formula from [18] for the number of linearized multi de Bruijn sequences starting with k-mer y:

$$|\mathcal{LC}_y((m),(q),k)| = \frac{1}{q^k}\left(\frac{(mq)!}{(m!)^q}\right)^{q^{k-1}}. \tag{47}$$

Table 2: Parameters for (a) classical de Bruijn sequences, in which a string over an alphabet of size q has all k-mers exactly once, and (b) multi de Bruijn sequences, with each k-mer occurring m times. Parameters m, q, k are positive integers.

Input parameters	**(a) Classical de Bruijn**	**(b) Multi de Bruijn**
# phases	$p = 1$	$p = 1$
k-mer multiplicity per phase	$\vec{m} = (1)$	$\vec{m} = (m)$
Alphabet size per phase	$\vec{q} = (q)$	$\vec{q} = (q)$
Word size	$k > 0$	$k > 0$
Derived parameters		
Vertex degrees per phase	$\vec{d} = (q)$	$\vec{d} = (mq)$
# vertices per phase	$\vec{n} = (q^{k-1})$	$\vec{n} = (q^{k-1})$
# k-mers per phase	$\vec{\theta} = (q^k)$	$\vec{\theta} = (q^k)$
Sequence length	$\ell = q^k$	$\ell = mq^k$

4.3 Counting cyclic sequences

In [18], we counted cyclic multi de Bruijn sequences that start with a given k-mer y, accounting for possible rotational symmetries; also see Van Aardenne-Ehrenfest and de Bruijn [20, §5]. Now we use a similar argument for phased multi de Bruijn sequences.

Suppose cyclic sequence $(s) \in \mathcal{C}(\vec{m}, \vec{q}, k)$ decomposes as $s = t^d$. For each $i \in \mathbb{P}$, each $y \in K_i$ occurs m_i/d times in t, so $(t) \in \mathcal{C}(\vec{m}/d, \vec{q}, k)$, where $\vec{m}/d = (m_0/d, \ldots, m_{p-1}/d)$. This requires that $d|m_i$ for all i. Equivalently, it requires $d|g$ where

$$g = \gcd(m_0, m_1, \ldots, m_{p-1}) \ . \tag{48}$$

For a cyclic sequence (s) or linearized sequence s, recall that its *rotational order* is the largest d for which s can be written as $s = t^d$. Define $\mathcal{C}^{(d)}(\vec{m}, \vec{q}, k)$ as the order d sequences in $\mathcal{C}(\vec{m}, \vec{q}, k)$, and define $\mathcal{LC}^{(d)}(\vec{m}, \vec{q}, k)$ and $\mathcal{LC}_y^{(d)}(\vec{m}, \vec{q}, k)$ similarly. Now decompose the sets of linearized and cyclic phased multi de Bruijn sequences into disjoint parts by the rotational order of each sequence:

$$\text{(a) } \mathcal{LC}_y(\vec{m}, \vec{q}, k) = \bigcup_{d|g} \mathcal{LC}_y^{(d)}(\vec{m}, \vec{q}, k) \qquad \text{(b) } \mathcal{C}(\vec{m}, \vec{q}, k) = \bigcup_{d|g} \mathcal{C}^{(d)}(\vec{m}, \vec{q}, k). \tag{49}$$

Lemma 16. *Let $i \in \mathbb{P}$ be any phase, $y \in K_i$ be a k-mer, and $d|g$. Then*

(a) $|\mathcal{LC}_y^{(d)}(\vec{m}, \vec{q}, k)| = |\mathcal{LC}_y^{(1)}(\vec{m}/d, \vec{q}, k)|$

(b) $|\mathcal{LC}_y^{(d)}(\vec{m}, \vec{q}, k)| = (m_i/d)|\mathcal{C}^{(d)}(\vec{m}, \vec{q}, k)|$

(c) $|\mathcal{C}^{(d)}(\vec{m}, \vec{q}, k)| = |\mathcal{C}^{(1)}(\vec{m}/d, \vec{q}, k)|$.

Proof. (a) We give a bijection $\psi : \mathcal{LC}_y^{(d)}(\vec{m},\vec{q},k) \to \mathcal{LC}_y^{(1)}(\vec{m}/d,\vec{q},k)$.

Each $s \in \mathcal{LC}^{(d)}(\vec{m},\vec{q},k)$ can be written $s = t^d$; set $\psi(s) = t$. Since s has order d, then t has order $d/d = 1$.

Both s and t begin with the same k-mer, y. (Technicality: if s or t has length below k, recall that we extract a k-mer by wrapping around the sequence as many times as needed, reusing sequence entries. The initial k-mers of t and t^d are the same, even if we need to wrap around the sequence.)

Each phase j k-mer occurs m_j times in s and m_j/d times in t, so $t \in \mathcal{LC}_y^{(1)}(\vec{m}/d,\vec{q},k)$.

Conversely, let $t \in \mathcal{LC}_y^{(1)}(\vec{m}/d,\vec{q},k)$. Then t^d has order d and $d \cdot (m_j/d) = m_j$ occurrences of each phase j k-mer, so $t^d \in \mathcal{LC}_y^{(d)}(\vec{m},\vec{q},k)$. Thus, ψ is a bijection, proving (a).

(b) Let $s \in \mathcal{LC}_y^{(d)}(\vec{m},\vec{q},k)$. Consider the m_i rotations of s that start with y. They all give the same cycle (s) in $\mathcal{C}(\vec{m},\vec{q},k)$. The first m_i/d rotations give distinct linearizations, while the other rotations repeat these same linearizations a total of d times each.

(c) Extend ψ to cyclic sequences, $\psi : \mathcal{C}^{(d)}(\vec{m},\vec{q},k) \to \mathcal{C}^{(1)}(\vec{m}/d,\vec{q},k)$. For $(s) \in \mathcal{C}^{(d)}(\vec{m},\vec{q},k)$, set $\psi((s)) = (t)$, where $s = t^d$. Conversely, $(t) \in \mathcal{C}^{(1)}(\vec{m}/d,\vec{q},k)$ has inverse (t^d). The analysis of the multiplicities is the same as in (a).

Since (s) has multiple linearizations, we must show that ψ on cyclic sequences is well-defined. Using linearization $S = \rho^h(s)$ instead of s gives $S = T^d$, where $T = \rho^h(t)$, and would give $\psi((S)) = (T)$. Since $(T) = (t)$, this map is well-defined. A similar argument shows the inverse is well-defined. □

Let $\vec{m}' = \vec{m}/g$. Note that $\gcd(m'_0,\ldots,m'_{p-1}) = g/g = 1$. For $d \in \mathbb{Z}^+$, define

$$f(d) = |\mathcal{LC}_y(d\,\vec{m}',\vec{q},k)| = W(d\,\vec{m}',\vec{q},k)/\theta_i \qquad h(d) = |\mathcal{LC}_y^{(1)}(d\,\vec{m}',\vec{q},k)|. \tag{50}$$

Note by Lemma 6 that parameters $(d\,\vec{m}',\vec{q},k)$ are compatible iff $(\vec{m},\vec{q},k)$ are compatible. We assume $(\vec{m},\vec{q},k)$ are compatible (if not, then $f(d) = h(d) = 0$). Then

$$\begin{aligned} &|\mathcal{LC}_y(\vec{m},\vec{q},k)| = |\mathcal{LC}_y(g\,\vec{m}',\vec{q},k)| = f(g) \\ \text{and}\quad &|\mathcal{LC}_y^{(d)}(\vec{m},\vec{q},k)| = |\mathcal{LC}_y^{(1)}(\vec{m}/d,\vec{q},k)| = |\mathcal{LC}_y^{(1)}((g/d)\,\vec{m}',\vec{q},k)| = h(g/d). \end{aligned}$$

Using (49a), $|\mathcal{LC}_y(\vec{m},\vec{q},k)| = \sum_{d|g} |\mathcal{LC}_y^{(d)}(\vec{m},\vec{q},k)|$, or equivalently, $f(g) = \sum_{d|g} h(g/d) = \sum_{d|g} h(d)$. By Möbius inversion, $h(g) = \sum_{d|g} \mu(d) f(g/d)$, where $\mu(n)$ is the Möbius function. By (42), this is

$$|\mathcal{LC}_y^{(1)}(\vec{m},\vec{q},k)| = h(g) = \sum_{d|g} \mu(d) \cdot |\mathcal{LC}_y(\vec{m}/d,\vec{q},k)| = \frac{1}{\theta_i} \sum_{d|g} \mu(d) \cdot W(\vec{m}/d,\vec{q},k). \tag{51}$$

Thus, by Lemma 16(a,b),

$$|\mathcal{LC}_y^{(d)}(\vec{m},\vec{q},k)| = h(g/d) = \frac{1}{\theta_i} \sum_{d' \,|\, (g/d)} \mu(d') \cdot W\left(\frac{\vec{m}}{d\,d'},\vec{q},k\right) \tag{52}$$

$$|\mathcal{C}^{(d)}(\vec{m},\vec{q},k)| = \frac{d}{m_i} \cdot |\mathcal{LC}_y^{(d)}(\vec{m},\vec{q},k)| = \frac{p\,d}{\ell} \sum_{d' \,|\, (g/d)} \mu(d') \cdot W\left(\frac{\vec{m}}{d\,d'},\vec{q},k\right). \tag{53}$$

Note (11) gives $m_i\theta_i = \ell/p$, so $(d/m_i)(1/\theta_i) = pd/\ell$. Now combine (53) with (49b):

$$|\mathcal{C}(\vec{m},\vec{q},k)| = \sum_{d|g} |\mathcal{C}^{(d)}(\vec{m},\vec{q},k)| = \frac{p}{\ell}\sum_{d|g} d \cdot \sum_{d' \mid (g/d)} \mu(d') \cdot W\left(\frac{\vec{m}}{d\,d'},\vec{q},k\right).$$

Set $r = d\,d'$, so $d' = r/d$:

$$= \frac{p}{\ell}\sum_{r|g}\sum_{d|r} d \cdot \mu(r/d) \cdot W(\vec{m}/r,\vec{q},k) = \frac{p}{\ell}\sum_{r|g}\phi(r)\cdot W(\vec{m}/r,\vec{q},k)\,, \tag{54}$$

where $\phi(r) = \sum_{d|r} d \cdot \mu(r/d)$ is the Euler totient function.

5 Example: shifting suits on cards

Consider a deck of cards $\mathcal{A} \times \mathcal{B}$, where $\mathcal{A}$ and $\mathcal{B}$ are the sets of denominations and suits, with sizes $A = |\mathcal{A}|$ and $B = |\mathcal{B}|$. Represent each card as ab, with $a \in \mathcal{A}$ and $b \in \mathcal{B}$. Let Shuffles be the set of multiset permutations of m indistinguishable decks shuffled together; such a permutation has the form

$$s = a_0b_0,\ a_1b_1,\ a_2b_2,\ \ldots,\ a_{L-1}b_{L-1} \tag{55}$$

where $a_i \in \mathcal{A}$, $b_i \in \mathcal{B}$, $L = mAB$, and every distinct card occurs m times. The number of distinguishable shuffles is $|\,\textsc{Shuffles}\,| = (mAB)!/m!^{AB}$.

We can view s as a linearized de Bruijn sequence with one phase, over an alphabet $\Omega_0 = \mathcal{A} \times \mathcal{B}$ of size $q_0 = AB$, with every $k = 1$-mer over Ω_0 occurring exactly m times; see Fig. 1(a). The number of such sequences is found by (44) and gives the same result:

$$|\,\textsc{Shuffles}\,| = |\mathcal{LC}_{[[0]]}((m),(AB),1)| = \frac{(mAB)!}{m!^{AB}}\,. \tag{56}$$

We define an operation that shifts each suit one card to the right (with the rightmost card's suit wrapping around to the start): for s as defined in (55), set

$$\textsc{ShiftSuit}(s) = a_0b_{L-1},\ a_1b_0,\ a_2b_1,\ \ldots,\ a_{L-1}b_{L-2} \tag{57}$$

and for a cyclic sequence (s), set $\textsc{ShiftSuit}((s)) = (\textsc{ShiftSuit}(s))$. Now, $\textsc{ShiftSuit}(s)$ may or may not be a permutation of the m decks (see Fig. 1(d,e)); we will determine the probability that it is. We say that s is *shiftable* if the multisets of cards in both s and $\textsc{ShiftSuit}(s)$ are the same (equivalently, if $\textsc{ShiftSuit}(s)$ is also a permutation of m decks of cards).

Next, we form the *two-phase representation*, with each card represented as two characters: denomination in phase 0 and suit in phase 1. This gives a linearized phased string of length $\ell = 2L$. Fig. 1(b) has $s = 2♣4♢4♣3♡2♡4♡3♢3♣2♢$, a sequence of length 18. Reading off the 2-mers at every phase 0 position gives the initial shuffle

$$2♣, 4♢, 4♣, 3♡, 2♡, 4♡, 3♢, 3♣, 2♢.$$

Denominations	$\mathcal{A} = \{2, 3, 4\}$
Suits	$\mathcal{B} = \{\clubsuit, \diamondsuit, \heartsuit\}$
Deck	$\mathcal{A} \times \mathcal{B} = \{2\clubsuit, 2\diamondsuit, 2\heartsuit, 3\clubsuit, 3\diamondsuit, 3\heartsuit, 4\clubsuit, 4\diamondsuit, 4\heartsuit\}$

(a) One-phase representation of a particular shuffle
Sequence of length 9; entries from $\boldsymbol{\mathcal{A} \times \mathcal{B}}$

$$2\clubsuit,\ 4\diamondsuit,\ 4\clubsuit,\ 3\heartsuit,\ 2\heartsuit,\ 4\heartsuit,\ 3\diamondsuit,\ 3\clubsuit,\ 2\diamondsuit$$

(b) Two-phase representation of same shuffle
Sequence of length 18; entries alternate from $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$

$$2,\ \clubsuit,\ 4,\ \diamondsuit,\ 4,\ \clubsuit,\ 3,\ \heartsuit,\ 2,\ \heartsuit,\ 4,\ \heartsuit,\ 3,\ \diamondsuit,\ 3,\ \clubsuit,\ 2,\ \diamondsuit$$

(c) Parallel lines representation of same shuffle
Two sequences of length 9, one over $\boldsymbol{\mathcal{A}}$ and the other over $\boldsymbol{\mathcal{B}}$

Denominations	2,	4,	4,	3,	2,	4,	3,	3,	2
Suits	$\clubsuit$,	$\diamondsuit$,	$\clubsuit$,	$\heartsuit$,	$\heartsuit$,	$\heartsuit$,	$\diamondsuit$,	$\clubsuit$,	$\diamondsuit$

(d) Shiftable permutation

Shuffle s	$2\clubsuit,\ 4\diamondsuit,\ 4\clubsuit,\ 3\heartsuit,\ 2\heartsuit,\ 4\heartsuit,\ 3\diamondsuit,\ 3\clubsuit,\ 2\diamondsuit$
ShiftSuit(s)	$2\diamondsuit,\ 4\clubsuit,\ 4\diamondsuit,\ 3\clubsuit,\ 2\heartsuit,\ 4\heartsuit,\ 3\heartsuit,\ 3\diamondsuit,\ 2\clubsuit$

(e) Non-shiftable permutation

Shuffle t	$2\clubsuit,\ 4\diamondsuit,\ 3\clubsuit,\ 4\clubsuit,\ 3\heartsuit,\ 2\heartsuit,\ 4\heartsuit,\ 3\diamondsuit,\ 2\diamondsuit$
ShiftSuit(t)	$2\diamondsuit,\ 4\clubsuit,\ 3\diamondsuit,\ 4\clubsuit,\ 3\clubsuit,\ 2\heartsuit,\ 4\heartsuit,\ 3\heartsuit,\ 2\diamondsuit$

Figure 1: Representations of a shuffled deck of cards. We use a 9-card deck, shown at the top. The same shuffle is represented in three ways: (a) *one-phase representation:* a sequence of 9 cards; (b) *two-phase representation:* a sequence of length 18, with entries alternating from denominations and suits; and (c) *parallel lines representation:* a length 9 sequence of denominations, and a length 9 sequence of suits. In (d)–(e), we illustrate the ShiftSuit operation, which cyclically shifts suits one card to the right (wrapping around the end). In (d), this operation preserves the multiset of cards, so the shuffle is *shiftable*. But a different shuffle in (e) is not shiftable, as the multiset of cards changes (shifting yields two $2\diamondsuit$'s and two $4\clubsuit$'s, but no $2\clubsuit$'s or $4\diamondsuit$'s).

Reading off the 2-mers at every phase 1 position (starting with the final character of s and wrapping around) effectively shifts the suits by one card, but represents each card as suit and then denomination:

$$\diamondsuit 2, \clubsuit 4, \diamondsuit 4, \clubsuit 3, \heartsuit 2, \heartsuit 4, \heartsuit 3, \diamondsuit 3, \clubsuit 2.$$

Rewriting this as denomination first, suit second, gives

$$\textsc{ShiftSuit}(s) = 2\diamondsuit, 4\clubsuit, 4\diamondsuit, 3\clubsuit, 2\heartsuit, 4\heartsuit, 3\heartsuit, 3\diamondsuit, 2\clubsuit$$

(see Fig. 1(d)). This s is shiftable; see Fig. 1(e) for an example that is not shiftable.

Parameters for the general cases of one- and two-phase representations are shown in Table **??**. The number of linearized shiftable sequences is found using the two-phase representation: it's the number of linearized two-phase sequences in phase 0 (before shifting) or phase 1 (after shifting). By (44):

$$\left|\mathcal{LC}_{[[0]]}\big((m,m),(A,B),2\big)\right| = \left(\frac{(mA)!}{(m!)^A}\right)^B \cdot \left(\frac{(mB)!}{(m!)^B}\right)^A = \frac{(mA)!^B(mB)!^A}{(m!)^{2AB}} \,. \tag{58}$$

Divide by $|\,\textsc{Shuffles}\,|$ (Eq. (56)) to get the probability that a random permutation of m decks is shiftable:

$$\frac{(mA)!^B\,(mB)!^A}{(m!)^{2AB}} \Big/ \frac{(mAB)!}{(m!)^{AB}} = \frac{(mA)!^B\,(mB)!^A}{(mAB)!\,(m!)^{AB}} \,. \tag{59}$$

The probability a random permutation of one deck is shiftable is the special case $m = 1$:

$$\frac{(A!)^B \cdot (B!)^A}{(AB)!} \,. \tag{60}$$

Next, we consider m decks shuffled together to form a cyclic sequence (instead of linearized). Again we count the total number of shuffles, and how many are shiftable.

Counting all shuffles of m decks in cyclic case: Use the one phase representation, with parameters in Table **??**(a). Shuffling m decks to get a cyclic ordering gives a cyclic multi de Bruijn sequence. Evaluate the number of cyclic shuffles shuffles using (54) (or the simpler version of this formula for multi de Bruijn sequences in [18]). Note that $p/\ell = 2/(2mAB) = 1/(mAB)$:

$$|\mathcal{C}((m),(AB),1)| = \frac{1}{mAB} \sum_{r|m} \phi(r) \cdot \frac{(mAB/r)!}{(m/r)!^{AB}} \,. \tag{61}$$

Counting shiftable shuffles of m decks in cyclic case: Use the two phase representation, with parameters in Table **??**(b), and compute (54):

$$|\mathcal{C}((m,m),(A,B),2)| = \frac{1}{mAB} \sum_{r|m} \phi(r) \cdot \frac{(mA/r)!^B\,(mB/r)!^A}{(m/r)!^{2AB}}. \tag{62}$$

Divide (62) by (61) to get the probability that a cyclic shuffle of m decks is shiftable:

$$\left(\sum_{r|m} \phi(r) \cdot \frac{(mA/r)!^B \, (mB/r)!^A}{(m/r)!^{2AB}} \right) \Bigg/ \left(\sum_{r|m} \phi(r) \cdot \frac{(mAB/r)!}{(m/r)!^{AB}} \right). \tag{63}$$

For one deck ($m = 1$), each sum has only one term ($r = 1$), with $\phi(1) = 1$ and $m/r = 1/1 = 1$, so this probability simplifies to (60).

Note: While ShiftSuit is not physically practical with ordinary cards, we can take two adjacent lines or circles of objects, with one line/circle of mAB objects labelled by denominations, and the other labelled by suits; see Fig. 1(c). Consider a partner dance formation with two lines or circles of people (mAB in each line) sequentially switching partners. Each person in the first line has an outfit selected from set $\mathcal{A}$, and each person in the second line has an outfit selected from $\mathcal{B}$. In the first line, mB people have each outfit from $\mathcal{A}$, and in the second line, mA people have each outfit from $\mathcal{B}$. Valid pairings of the lines yield each pair in $\mathcal{A} \times \mathcal{B}$ exactly m times. Shifting suits in the cards corresponds to everyone in the second line advancing to the next partner. A configuration is shiftable when this shift preserves the how many partners have each combination of outfits.

Relation to other work: By (38), the number of spanning trees of the directed graph $G_m = G((m, m), (A, B), 2)$ rooted at any specified vertex is

$$\frac{p}{\ell} {d_0}^{n_0} {d_1}^{n_1} = \frac{2}{2mAB} (mB)^A (mA)^B = m^{A+B-1} B^{A-1} A^{B-1}. \tag{64}$$

Let $\mathbf{K}_{A,B}$ denote the complete bipartite graph; we use boldface to distinguish it from the sets of k-mers, K_i. The graph $G_1 = G((1, 1), (A, B), 2)$ is essentially equivalent to $\mathbf{K}_{A,B}$: each edge $\{v, w\}$ of undirected graph $\mathbf{K}_{A,B}$ corresponds to a pair of directed edges (v, w) and (w, v) in G_1, so the adjacency matrices of G_1 and $\mathbf{K}_{A,B}$ are identical. In 1962, Scoins [16] proved (in different notation) that the number of spanning trees of $\mathbf{K}_{A,B}$ is $A^{B-1}B^{A-1}$ (also see Hartsfield and Werth [9], and an extension to complete multipartite graphs by Onodera [13]). This agrees with (64) for $m = 1$.

For $m > 1$, graph $\mathbf{K}_{A,B}$ has $A+B$ vertices and each of its spanning trees has $A+B-1$ edges. Form a new undirected multigraph H from $\mathbf{K}_{A,B}$ by giving each edge multiplicity m. Each spanning tree T of $\mathbf{K}_{A,B}$ yields m^{A+B-1} spanning trees of H (for each edge of T, choose any of the m corresponding edges in H) so the number of spanning trees of H is the rightmost side of (64).

6 Example: Markov chains and universal cycles

In this section, we show that phased sequences represent walks through a periodic irreducible Markov chain, and phased multi de Bruijn sequences are *universal cycles* encoding all walks of a certain length in such Markov chains.

Chung et al. [2] introduced *universal cycles*, a generalization of de Bruijn sequences; also see Hurlbert [11]. These are circular sequences encoding a set T of combinatorial

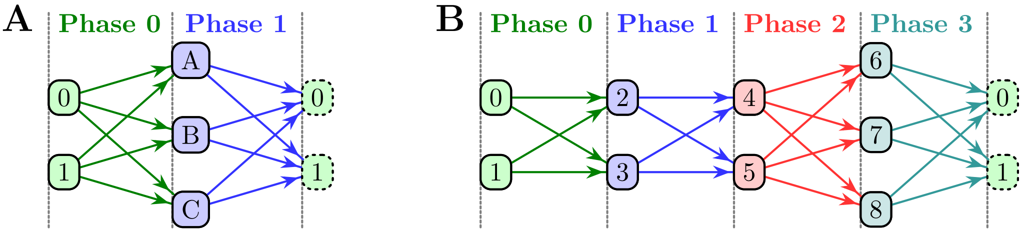

Figure 2: Periodic irreducible Markov chains. (A) $G((1,1),(2,3),1)$ for Example 1. (B) $G((1,1,1,1),(2,2,2,3),1)$ for Appendix A.2. The phase 0 vertices (0 and 1 in both examples) are each drawn twice (solid border on the left, dotted on the right), and the two copies are identified with each other.

objects, such as all permutations of a set or all partitions of a set. Let $k \geqslant 1$, and specify a rule for how k-mers over an alphabet Ω represent elements of T. Let (s) be a circular sequence over Ω. If every element of T is represented exactly once among the k-mers in (s), then (s) is a *universal cycle* for T.

Let p and $q_0, \ldots, q_{p-1}$ be positive integers, and $\vec{1} = (1, \ldots, 1)$ be p ones. The directed graph $H = G(\vec{1}, \vec{q}, 1)$ is a finite state automaton for phased strings with alphabet sizes $\vec{q}$. Assigning positive probabilities to the edges (with outgoing probabilities on each vertex summing to 1) gives an irreducible Markov chain in which all vertices have period p. Deleting certain edges (or setting their probabilities to 0) may also give an irreducible Markov chain with all vertices having period p, but we are not treating that case here.

Let $k \geqslant 1$ with $p|k$. Lemma 5 gives that $(\vec{1}, \vec{q}, k)$ are compatible parameters.

Label walks in H by their vertex sequence. Phased strings in $\mathcal{S}_{i,r}$ represent walks with $r-1$ edges in H, starting in phase i. E.g., in Example 1, the phased string $B0C1A1$ (length 6 characters) gives a walk in Fig. 2(A) (length 5 edges). Similarly, each length $k-1$ walk in H is represented by a phased k-mer. Each phased multi de Bruijn sequence $(s) \in \mathcal{C}(\vec{1}, \vec{q}, k)$ gives a universal cycle for all length $k-1$ walks in H: such sequences have every phased k-mer exactly once, and hence, every length $k-1$ walk in H exactly once.

Various authors have studied periodic irreducible Markov chains. Çinlar [3], Bonhoure et al. [1], and Kirkland [12], each studied properties of periodic irreducible Markov chains in general; in our notation, these are Markov chains on certain subgraphs of $G(\vec{1}, \vec{q}, 1)$. Pajarinen and Peltonen [14] studied "periodic finite state controllers"; in our notation, these are Markov chains on $G(\vec{1}, q\vec{1}, 1)$.

7 Change of basis to simplify the Laplacian matrix

In this section, we use a change of basis to simplify the Laplacian matrix, $L = D - A$, and compute its characteristic polynomial. This gives a new proof for enumerating classical de Bruijn sequences ($p = 1$, $\vec{m} = (1)$): the prior solution methods were to induct on k with a graph doubling construction (Sainte-Marie [15], de Bruijn [5], and van Aardenne-Ehrenfest and de Bruijn [20]), or to find eigenvalues of A^{k-1} (Dawson and Good [4]).

Throughout this section, we assume $k \geqslant 2$; we handle $k = 1$ at the end of this section (Theorem 27) because matrix A has different formulas (24)–(25) for $k \geqslant 2$ vs. $k = 1$.

7.1 Canonical alphabet

Let $(\vec{m}, \vec{q}, k)$ be compatible parameters. We define alphabets $\Omega_0, \ldots, \Omega_{p-1} \subseteq \mathbb{Z}$ as follows. For $i \in \mathbb{P}$, set

$$\begin{aligned} \alpha_i &= q_0 + \cdots + q_{i-1} & \Omega_i &= \{\alpha_i, \alpha_i + 1, \ldots, \zeta_i\} \\ \zeta_i &= \alpha_i + q_i - 1 & \Omega &= \Omega_0 \cup \cdots \cup \Omega_{p-1} = \{0, \ldots, \zeta_{p-1}\} . \end{aligned} \tag{65}$$

Order the alphabet Ω as integers, and order $\mathcal{S}_{i,r}$ lexicographically as sequences over Ω. Other alphabets may be used by forming bijections with these alphabets; e.g., Example 1 used $\Omega_0 = \{0, 1\}$ (good as-is) and $\Omega_1 = \{A, B, C\}$ (corresponding here to $\{2, 3, 4\}$), ordered as $0 < 1 < A < B < C$.

The minimum and maximum vertices in phase $i \in \mathbb{P}$ are:

$$O_i = \alpha_i \alpha_{i+1} \ldots \alpha_{i+k-2} \qquad\qquad Z_i = \zeta_i \zeta_{i+1} \ldots \zeta_{i+k-2} \tag{66}$$

and the minimum and maximum vertices in V are O_0 and Z_{p-1}. More generally, the maximum string in $\mathcal{S}_{i,r}$ is

$$Z_{i,r} = \zeta_i \zeta_{i+1} \ldots \zeta_{i+r-1} . \tag{67}$$

For a phased string $s = s_0 s_1 \ldots s_{r-1} \in \mathcal{S}_{i,r}$, set

$$\begin{aligned} \mathrm{ZL{\small EN}}(s) &= \max\{h \in \{0, \ldots, r\} : \mathrm{P{\small RE}}_h(s) = Z_{i,h}\} \\ &= \max\{h \in \{0, \ldots, r\} : s_0 s_1 \ldots s_{h-1} = \zeta_i \zeta_{i+1} \ldots \zeta_{i+h-1}\} . \end{aligned} \tag{68}$$

That is, $h = \mathrm{ZL{\small EN}}(s)$ is the length of the longest prefix of s comprised of maximum characters in each phase; the next character either isn't maximum, or $h = r$ so this prefix is all of s. Next, set

$$Z^+(s) = \begin{cases} s_0 s_1 \ldots s_{h-1} \zeta_{i+h} s_{h+1} \ldots s_{r-1} & \text{if } \mathrm{ZL{\small EN}}(s) < r \\ \text{(where } h = \mathrm{ZL{\small EN}}(s)); & \\ \text{“undefined”} & \text{if } \mathrm{ZL{\small EN}}(s) = r. \end{cases} \tag{69}$$

Example 17. Let $\vec{q} = (2, 2, 3)$, $\Omega_0 = \{0, 1\}$, $\Omega_1 = \{2, 3\}$, $\Omega_2 = \{4, 5, 6\}$, and $k = 5$.

- Max characters by phase: $\zeta_0 = 1$, $\zeta_1 = 3$, and $\zeta_2 = 6$
- Max vertices by phase: $Z_0 = 1361$, $Z_1 = 3613$, and $Z_2 = 6136$
  Each has $\mathrm{ZL{\small EN}}(Z_i) = 4$ and $Z^+(Z_i)$ undefined
- For $s = 61\underline{2}4$: $\mathrm{ZL{\small EN}}(s) = 2$ and $Z^+(s) = 61\underline{3}4$
  (the underlined position changed)

We'll use ZLen and Z^+ on strings of length $r = k-1$ (vertices) and $r = k-2$ (vertex with first or last character removed). These functions have the following properties:

Lemma 18. *Let $s = s_0 \ldots s_{r-1} \in \mathcal{S}_{i,r}$, with $i \in \mathbb{P}$ and $r \geqslant 1$. Then*

(a) $\quad Z^+(s) > s \qquad$ *provided* $\mathrm{ZLen}(s) \leqslant r-1$.

(b) *If $s_0 = \zeta_i$, then*

$$\begin{aligned} \mathrm{ZLen}(\mathrm{Suf}(s)) &= \mathrm{ZLen}(s) - 1 && \textit{(no restriction)} \\ Z^+(\mathrm{Suf}(s)) &= \mathrm{Suf}(Z^+(s)) && \textit{provided } \mathrm{ZLen}(s) \leqslant r-1. \end{aligned}$$

(c) *If $r \geqslant 2$, then*

$$\begin{aligned} \mathrm{ZLen}(\mathrm{Pre}(s)) &= \mathrm{ZLen}(s) && \textit{provided } \mathrm{ZLen}(s) \leqslant r-1 \\ Z^+(\mathrm{Pre}(s)) &= \mathrm{Pre}(Z^+(s)) && \textit{provided } \mathrm{ZLen}(s) \leqslant r-2. \end{aligned}$$

(d) *Let $t \in \mathcal{S}_{i,r}$. Then $s \leqslant t$ implies* $\mathrm{ZLen}(s) \leqslant \mathrm{ZLen}(t)$.

Proof. (a) If $\mathrm{ZLen}(s) \leqslant r-1$, then $Z^+(s)$ replaces a single entry of s by a larger value, giving $Z^+(s) > s$; while if $\mathrm{ZLen}(s) = r$, then $Z^+(s)$ is undefined.

(b) If $s_0 = \zeta_0$, then $\mathrm{ZLen}(s) \geqslant 1$, and for each $h \in \{1, \ldots, r\}$, we have

$$\underbrace{s_0 s_1 \ldots s_{h-1}}_{\mathrm{Pre}_h(s)} = \zeta_i \zeta_{i+1} \ldots \zeta_{i+h-1} \quad \text{iff} \quad \underbrace{s_1 \ldots s_{h-1}}_{\mathrm{Pre}_{h-1}(\mathrm{Suf}(s))} = \zeta_{i+1} \ldots \zeta_{i+h-1},$$

so $\mathrm{ZLen}(\mathrm{Suf}(s)) = \mathrm{ZLen}(s) - 1$. (This doesn't hold if $s_0 \neq \zeta_0$, because then $\mathrm{ZLen}(s) = 0$ while $0 \leqslant \mathrm{ZLen}(\mathrm{Suf}(s)) \leqslant r-1$.)

Set $h = \mathrm{ZLen}(s) \geqslant 1$, and further assume $\mathrm{ZLen}(s) \leqslant r-1$, so $1 \leqslant h \leqslant r-1$. Then

$$Z^+(\mathrm{Suf}(s)) = \mathrm{Suf}(Z^+(s)) = s_1 \ldots s_{h-1} \zeta_{i+h} s_{h+1} \ldots s_{r-1}\,.$$

(c) Since $\mathrm{ZLen}(s) \leqslant r-1$, both $\mathrm{ZLen}(s)$ and $\mathrm{ZLen}(\mathrm{Pre}(s))$ take the maximum of the same set, so they're equal. Set $h = \mathrm{ZLen}(s) = \mathrm{ZLen}(\mathrm{Pre}(s))$ and further assume $h \leqslant r-2$. Then

$$Z^+(\mathrm{Pre}(s)) = \mathrm{Pre}(Z^+(s)) = s_0 \ldots s_{h-1} \zeta_{i+h} s_{h+1} \ldots s_{r-2}\,.$$

We required $r \geqslant 2$ and $h \leqslant r-2$, so this is well-defined.

(d) Let $h = \mathrm{ZLen}(s)$; then $\mathrm{Pre}_h(s) = Z_{i,h}$. As $s \leqslant t$, then $Z_{i,h} = \mathrm{Pre}_h(s) \leqslant \mathrm{Pre}_h(t)$; but $Z_{i,h}$ is the maximum element of $\mathcal{S}_{i,h}$, and $\mathrm{Pre}_h(t) \in \mathcal{S}_{i,h}$, so $\mathrm{Pre}_h(t) = Z_{i,h}$. Thus, $\mathrm{ZLen}(t) \geqslant h$.

Note about restrictions on $\mathrm{ZLen}(s)$*:* If $\mathrm{ZLen}(s) = r$, then $s = Z_{i,r}$ (maximum element in $\mathcal{S}_{i,r}$), while $Z^+(s)$ and $Z^+(\mathrm{Suf}(s))$ are undefined. Similarly, if $\mathrm{ZLen}(s) \geqslant r-1$, then $\mathrm{Pre}(s) = Z_{i,r-1}$ so $Z^+(\mathrm{Pre}(s))$ is undefined. In upcoming steps, we'll apply Z^+ to vertices ($r = k-1$); the maximum vertices (Z_i's) will be handled as special cases. □

7.2 Change of basis matrix

Define the matrix $R = (R_{u,v})$ indexed by $V \times V$ by:

$$R_{u,v} = \begin{cases} 1 & \text{if } v = u; \\ -1 & \text{if } \text{ZL}\textsc{en}(u) < k-1 \text{ and } v = Z^+(u); \\ 0 & \text{otherwise.} \end{cases} \tag{70}$$

Note that the first two cases are disjoint: v can't equal both u and $Z^+(u)$, since $u \neq Z^+(u)$ (by Lemma 18(a)). In the middle case, $\text{ZL}\textsc{en}(u) < k-1$ excludes the values $u = Z_i$ where $Z^+(u)$ is undefined.

We will use the following properties of R:

- Matrix R is upper-triangular since $R_{u,v} \neq 0$ requires $v = u$ or $v = Z^+(u) > u$.
- For $u \in \{Z_0, \ldots, Z_{p-1}\}$, there is only one nonzero entry on row u (namely $R_{u,u} = 1$). All other rows of R have exactly two nonzero entries.

We will show that the inverse matrix R^{-1} is:

$$(R^{-1})_{u,v} = \begin{cases} 1 & \text{if } \textsc{Phase}(u) = \textsc{Phase}(v) \\ & \text{and } \textsc{Suf}_{k-1-h}(u) = \textsc{Suf}_{k-1-h}(v), \text{ where } h = \text{ZL}\textsc{en}(v); \\ 0 & \text{otherwise.} \end{cases} \tag{71}$$

Note that R and R^{-1} depend on $\vec{q}$ and k, but not on $\vec{m}$. See Appendix A for two examples (with $p = 1$ and 4) of matrices R, R^{-1}, A, D, L, and changes of basis RAR^{-1} and RLR^{-1}.

Lemma 19. *The matrix R^{-1} given by (71) has the following properties:*

(a) It is upper triangular.

(b) For $u = Z_i$, the only nonzero entry on row u is the diagonal entry, $(R^{-1})_{u,u} = 1$.

Proof. (a) Suppose $(R^{-1})_{u,v} \neq 0$. Let $h = \text{ZL}\textsc{en}(v)$. By (71), both u and v have the same phase (call it i), and $\textsc{Suf}_{k-1-h}(u) = \textsc{Suf}_{k-1-h}(v)$. By (68), $\textsc{Pre}_h(v) = Z_{i,h}$, which is the maximum element in $\mathcal{S}_{i,h}$, so $\textsc{Pre}_h(u) \leqslant \textsc{Pre}_h(v)$. Combining the prefix and suffix gives $u \leqslant v$.

(b) By (a), we have $\textsc{Suf}_{k-1-h}(u) = \textsc{Suf}_{k-1-h}(v)$ and $\textsc{Pre}_h(v) = Z_{i,h}$; but for $u = Z_i$, we also have $\textsc{Pre}_h(u) = Z_{i,h}$, so $\textsc{Pre}_h(u) = \textsc{Pre}_h(v)$. Combining the prefix and suffix gives $v = u = Z_i$. □

While we used the notation R^{-1} for matrix (71), we must prove it is actually the inverse of R. We temporarily use the notation $T_{u,v}$ (instead of $(R^{-1})_{u,v}$) for matrix (71). We will show that $R \cdot T =$ identity, and thus, R and T are inverses.

Theorem 20. *Matrices (70) and (71) are inverses.*

Proof. For all $u, v \in V$, we will verify that the following holds for definitions (70)–(71):

$$(R \cdot T)_{u,w} = \sum_{v \in V} R_{u,v} \cdot T_{v,w} = \begin{cases} 1 & \text{if } u = w; \\ 0 & \text{otherwise.} \end{cases} \tag{72}$$

Both R and T are block diagonal matrices, partitioned by vertex phases. If u and w have different phases, then in every term of the sum in (72), one or both of $R_{u,v}$ and $T_{v,w}$ is 0, so $(R \cdot T)_{u,w} = 0$.

Since R and T are upper triangular, $R \cdot T$ has diagonal entries $(R \cdot T)_{u,u} = R_{u,u} T_{u,u} = 1 \cdot 1 = 1$.

Suppose $u = Z_i$ and $w \in V_i$ but $w \neq u$. The only nonzero entry in row u of R is $R_{u,u} = 1$, so $(R \cdot T)_{u,w} = R_{u,u} T_{u,w} = T_{u,w} = 0$ (by Lemma 19(b), as $u = Z_i$ but $w \neq u$).

The rest of this proof considers off-diagonal entries ($u \neq w$) in the same phase ($u, v \in V_i$), with $u \neq Z_i$. Since $u \neq Z_i$, then $\text{ZLen}(u) < k - 1$. By (70), row u of R has exactly two nonzero entries: $R_{u,u} = 1$ and $R_{u,x} = -1$ (where $x = Z^+(u)$). Thus:

$$(R \cdot T)_{u,w} = R_{u,u} T_{u,w} + R_{u,x} T_{x,w} = T_{u,w} - T_{x,w}. \tag{73}$$

Set $h_1 = \text{ZLen}(u)$ and $h_2 = \text{ZLen}(w)$. Split $u = u_0 \dots u_{k-1}$ into the first h_2 positions ($u' = \text{Pre}_{h_2}(u)$) and the remaining $k - 1 - h_2$ positions ($u'' = \text{Suf}_{k-1-h_2}(u)$):

$$u = u_0 \dots u_{k-1} = \underbrace{u_0 \dots u_{h_2 - 1}}_{u'} \; \underbrace{u_{h_2} u_{h_2+1} \dots u_{k-1}}_{u''} \; . \tag{74}$$

Split $x = x_0 \dots x_{k-1}$ into x' and x'', and $w = w_0 \dots w_{k-1}$ into w' and w'', in the same way (first h_2 positions vs. remaining $k - 1 - h_2$). In this notation, Eq. (71) gives:

$$T_{u,w} = \begin{cases} 1 & \text{if } u'' = w''; \\ 0 & \text{otherwise} \end{cases} \qquad\qquad T_{x,w} = \begin{cases} 1 & \text{if } x'' = w''; \\ 0 & \text{otherwise.} \end{cases} \tag{75}$$

By (69), u and $x = Z^+(u)$ differ in exactly one position: $u_{h_1} \neq x_{h_1}$. We'll show that the rightmost side of (73) is 0; the details depend on how h_1 and h_2 compare:

- *If $h_1 < h_2$:* Then $u' \neq x'$ and $u'' = x''$. Thus, either u'' and x'' both equal w'', or both don't equal w''. Either way, $T_{u,w} = T_{x,w}$, so by (73), $(R \cdot T)_{u,w} = T_{u,w} - T_{x,w} = 0$.
- *If $h_1 \geqslant h_2$:* Note that $\text{ZLen}(x) > \text{ZLen}(u)$ by Lemma 18(a,d), and $h_1 \geqslant h_2$ is $\text{ZLen}(u) \geqslant \text{ZLen}(w)$. Then $x' = u' = w' = Z_{i,h_2}$, but $x'' \neq w''$. Also, $u \neq w$ and $u' = w'$ gives $u'' \neq w''$. Thus, (73) gives $(R \cdot T)_{u,w} = 0 - 0 = 0$. ☐

Having proved Theorem 20, we resume using R^{-1} (instead of T) to denote matrix (71).

7.3 Laplacian matrix in new basis

Next, we change bases to simplify matrices A, D, and $L = D - A$:

$$\widetilde{A} = RAR^{-1} \qquad \widetilde{D} = RDR^{-1} = D \qquad \widetilde{L} = RLR^{-1} = D - \widetilde{A}\,. \tag{76}$$

Lemma 21. $RDR^{-1} = D$

Proof. Write R and D as block diagonal matrices, partitioned by vertex phases:

$$R = R_0 \oplus R_1 \oplus \cdots \oplus R_{p-1} \qquad D = D_0 \oplus D_1 \oplus \cdots \oplus D_{p-1}. \tag{77}$$

The i^{th} diagonal block of D is $D_i = d_i I$, where I is the identity matrix on indices V_i. Then $R_i\, D_i\, R_i{}^{-1} = R_i\,(d_i I)\, R_i{}^{-1} = d_i\,(R_i\, I\, R_i{}^{-1}) = d_i\, I = D_i$. ☐

Lemma 22. *Let $(\vec{m}, \vec{q}, k)$ be compatible parameters, with $k \geqslant 2$. Then $\widetilde{A} = RAR^{-1}$ is given as follows, where $u = u_0 \ldots u_{k-2} \in V_i$ and $v = v_0 \ldots v_{k-2} \in V_j$:*

$$\widetilde{A}_{u,v} = \begin{cases} m_i & \text{if } u_0 = \zeta_i,\ \textsc{Suf}(u) = \textsc{Pre}(v), \text{ and } v \neq Z_{i+1}; \\ d_i & \text{if } u_0 = \zeta_i,\ \textsc{Suf}(u) = \textsc{Pre}(v), \text{ and } v = Z_{i+1} \\ & \text{(equivalently, if } u = Z_i \text{ and } v = Z_{i+1}); \\ 0 & \text{otherwise.} \end{cases} \tag{78}$$

We will prove this shortly, after using it to compute $\widetilde{L}$. The case $k = 1$ is handled separately at the end of Sec. 7.4. On comparing $\widetilde{A}$ with the definition of A in (24):

- Rows with $u_0 \neq \zeta_i$ are zeroed out in $\widetilde{A}$.
- The bottom right corner of each nonzero block changes from $A_{Z_i,Z_{i+1}} = m_i$ to $\widetilde{A}_{Z_i,Z_{i+1}} = d_i$.
- All other positions are the same.

Lemma 23. *Let $(\vec{m}, \vec{q}, k)$ be compatible parameters, with $k \geqslant 2$. Then $\widetilde{L} = RLR^{-1}$ is given as follows, where $u = u_0 \ldots u_{k-2} \in V_i$ and $v = v_0 \ldots v_{k-2} \in V_j$:*

$$\text{For } p \geqslant 2\text{:} \quad \widetilde{L}_{u,v} = \begin{cases} d_i & \text{if } u = v; \\ -d_i & \text{if } u_0 = \zeta_i,\ \textsc{Suf}(u) = \textsc{Pre}(v), \text{ and } v = Z_{i+1} \\ & \text{(equivalently, if } u = Z_i \text{ and } v = Z_{i+1}); \\ -m_i & \text{if } u_0 = \zeta_i,\ \textsc{Suf}(u) = \textsc{Pre}(v), \text{ and } v \neq Z_{i+1}; \\ 0 & \text{otherwise.} \end{cases} \tag{79}$$

$$\text{For } p = 1\text{:} \quad \widetilde{L}_{u,v} = \begin{cases} d_0 & \text{if } u = v \neq Z_0; \\ 0 & \text{if } u = v = Z_0; \\ -m_0 & \text{if } u_0 = \zeta_0,\ \textsc{Suf}(u) = \textsc{Pre}(v), \text{ and } v \neq Z_0; \\ 0 & \text{otherwise.} \end{cases} \tag{80}$$

Proof. By (76), $\widetilde{L} = D - \widetilde{A}$.

By (78), any nonzero diagonal entry of $\widetilde{A}$ has $u = v$, $\textsc{Suf}(u) = \textsc{Pre}(v)$, and $u_0 = \zeta_i$; then $i = j$ and $j \equiv i + 1 \pmod p$, simultaneously. This can only happen for $p = 1$.

Thus, if $p \geqslant 2$, then $\widetilde{A}$ has no nonzero entries on the main diagonal. Then $\widetilde{L} = D - \widetilde{A}$ matches D on the diagonal (Eq. (30)) and $-\widetilde{A}$ off-diagonal (negative of (78)), yielding (79).

For $p = 1$, equations $u = v$ and $\textsc{Suf}(u) = \textsc{Pre}(v)$ are solved by $u = v = a^{k-1}$ for any $a \in \Omega_0$; combining with $u_0 = \zeta_0$, then $u = v = Z_0$ gives the only nonzero diagonal entry, $\widetilde{A}_{Z_0,Z_0} = d_0$. Then $\widetilde{L}_{Z_0,Z_0} = d_0 - d_0 = 0$. Other entries of $\widetilde{L}_{u,v}$ agree with the $p \geqslant 2$ case. We also simplify all phases to 0 (e.g., d_0 and Z_0 instead of d_i and Z_i), giving (80). ☐

Proof of Lemma 22. We'll show that $RA = \widetilde{A}R$, which is equivalent to $\widetilde{A} = RAR^{-1}$.

Let $u \in V_i$ and $w \in V$. If $w \notin V_{i+1}$, then $(RA)_{u,w} = (\widetilde{A}R)_{u,w} = 0$, due to the block partition by phase. For the rest of this proof, we assume $u \in V_i$ and $w \in V_{i+1}$. Write $u = u_0 \dots u_{k-2} \in V_i$ and $w = w_0 \dots w_{k-2} \in V_{i+1}$. We will evaluate the following sums, using the definitions of A, R, and $\widetilde{A}$ (Eqs. (24), (70), and (78)):

$$\text{(a) } (RA)_{u,w} = \sum_{v \in V_i} R_{u,v} \cdot A_{v,w} \qquad \text{(b) } (\widetilde{A}R)_{u,w} = \sum_{v \in V_{i+1}} \widetilde{A}_{u,v} \cdot R_{v,w} \,. \tag{81}$$

First we evaluate (81a). Each row of R has at most two nonzero terms, yielding

$$(RA)_{u,w} = \begin{cases} R_{u,u}A_{u,w} + R_{u,x}A_{x,w} = A_{u,w} - A_{x,w} & \text{if } u \neq Z_i; \\ R_{u,u}A_{u,w} = A_{u,w} & \text{if } u = Z_i, \end{cases} \tag{82}$$

where $x = Z^+(u)$ (provided $u \neq Z_i$). Recall that u and x only differ in one position. This leads to three cases, based on $h = \textsc{ZLen}(u)$:

$$\begin{aligned} &\text{(i) } h = 0 \text{ (equivalently, } u_0 \neq \zeta_0); \\ &\text{(ii) } 1 \leqslant h \leqslant k-2 \text{ (equivalently, } u_0 = \zeta_0 \text{ and } u \neq Z_i); \\ &\text{(iii) } h = k-1 \text{ (equivalently, } u = Z_i). \end{aligned} \tag{83}$$

Case (i): $u_0 \neq \zeta_i$: Then u and x differ only in position $h = 0$, so $\textsc{Suf}(x) = \textsc{Suf}(u)$. Thus, $A_{u,w} = A_{x,w}$ (both equal m_i or both equal 0), so $(RA)_{u,w} = A_{u,w} - A_{x,w} = 0$.

Case (ii): $u_0 = \zeta_i$ but $u \neq Z_i$: As u and x only differ in position h, and $h \geqslant 1$, then $\textsc{Suf}(x)$ and $\textsc{Suf}(u)$ contain this position, so $\textsc{Suf}(x) \neq \textsc{Suf}(u)$. Thus, $\textsc{Pre}(w)$ can equal at most one of $\textsc{Suf}(u)$ or $\textsc{Suf}(x)$ (but may not equal either), yielding:

For $u \in V_i$ in case (ii) and $w \in V_{i+1}$:

$$(RA)_{u,w} = A_{u,w} - A_{x,w} = \begin{cases} m_i - 0 = m_i & \text{if } \textsc{Pre}(w) = \textsc{Suf}(u); \\ 0 - m_i = -m_i & \text{if } \textsc{Pre}(w) = \textsc{Suf}(x); \\ 0 & \text{otherwise.} \end{cases} \tag{84}$$

Case (iii): $u = Z_i$: By (82), this case has $(RA)_{u,w} = A_{u,w}$. So for $u = Z_i$ and $w \in V_{i+1}$:

$$(RA)_{u,w} = \begin{cases} m_i & \text{if } \textsc{Pre}(w) = \textsc{Suf}(u); \\ 0 & \text{otherwise.} \end{cases} \tag{85}$$

Combining all cases and plugging in $x = Z^+(u)$ gives that for $u \in V_i$ and $w \in V$:

$$(RA)_{u,w} = \begin{cases} 0 & \text{if } u_0 \neq \zeta_i; \\ m_i & \text{if } u_0 = \zeta_i \text{ and } \textsc{Pre}(w) = \textsc{Suf}(u); \\ -m_i & \text{if } u_0 = \zeta_i \text{ and } u \neq Z_i \text{ and } \textsc{Pre}(w) = \textsc{Suf}(Z^+(u)); \\ 0 & \text{otherwise.} \end{cases} \tag{86}$$

Next, we'll evaluate $(\widetilde{A}R)_{u,w}$ and show that it equals $(RA)_{u,w}$. We can restrict sum (81b) to v satisfying the following; these are necessary conditions for nonzero terms, though not sufficient, so they may still yield terms that are 0:

$$\text{For } \widetilde{A}_{u,v} \neq 0: \quad v = u_1 \dots u_{k-2}\, c \quad \text{ for some } c \in \Omega_{i+k-1}. \tag{87}$$

$$\text{For } R_{v,w} \neq 0: \quad \text{either } v = w \text{ or } Z^+(v) = w. \tag{88}$$

We again set $h = \textsc{ZLen}(u)$ and get the same three cases, (i)–(iii) (see Eq. (83)):

Case (i): $u_0 \neq \zeta_i$: All $\widetilde{A}_{u,v} = 0$, so $(\widetilde{A}R)_{u,w} = 0$, agreeing with $(RA)_{u,w} = 0$ in case (i).

Case (ii): $u_0 = \zeta_i$ but $u \neq Z_i$: For v of form (87), $\textsc{ZLen}(v) = h - 1$. Note v and $Z^+(v)$ only differ in position $h - 1 \leqslant k - 3$, so their last position matches. Since $w = v$ or $Z^+(v)$ by (88), then w, v, and $Z^+(v)$ all match in their last position. Then (87) restricts to:

$$v = \ u_1 \dots u_{k-2}\, w_{k-2} \ \in V_{i+1}. \tag{89}$$

This reduces sum (81b) to just one term, $(\widetilde{A}R)_{u,w} = \widetilde{A}_{u,v} \cdot R_{v,w}$, with v given by (89). Note $v \neq Z_{i+1}$ (as $\textsc{ZLen}(v) < k - 1$), so $\widetilde{A}_{u,v} = m_i$. We evaluate $R_{v,w}$ for v of form (89):

- $R_{v,w} = 1$ iff $v = w$
  iff $\textsc{Pre}(v) = \textsc{Pre}(w)$ (since the last characters agree)
  iff $\textsc{Suf}(u) = \textsc{Pre}(w)$ (since $\textsc{Pre}(v) = \textsc{Suf}(u)$).

- $R_{v,w} = -1$ iff $Z^+(v) = w$
  iff $\textsc{Pre}(Z^+(v)) = \textsc{Pre}(w)$ (again, the last characters agree)
  iff $\textsc{Suf}(Z^+(u)) = \textsc{Pre}(w)$ (see details below).

  Note: We'll show $\textsc{Pre}(Z^+(v)) = Z^+(\textsc{Pre}(v)) = Z^+(\textsc{Suf}(u)) = \textsc{Suf}(Z^+(u))$:

  First, $\textsc{Pre}(Z^+(v)) = Z^+(\textsc{Pre}(v))$ by Lemma 18(c). In the lemma, $r = k - 1$. The lemma requires $r \geqslant 2$ (so $k - 1 \geqslant 2$) and $\textsc{ZLen}(v) \leqslant r - 2$ (so $h - 1 \leqslant k - 3$); we're in case (ii) ($1 \leqslant h \leqslant k - 2$), which implies both of those.

  Second, $\textsc{Pre}(v) = \textsc{Suf}(u)$ by (89), so $Z^+(\textsc{Pre}(v)) = Z^+(\textsc{Suf}(u))$.

  Finally, $Z^+(\textsc{Suf}(u)) = \textsc{Suf}(Z^+(u))$ by Lemma 18(b). This lemma requires $u_0 = \zeta_i$ and $\textsc{ZLen}(u) \leqslant r - 1$ (which is $h \leqslant k - 2$); both hold as we're in case (ii).

- $R_{v,w} = 0 \quad$ otherwise.

Thus, we have the following for $u \in V_i$ in case (ii) and $w \in V_{i+1}$:

$$(\widetilde{A}R)_{u,w} = \underbrace{\widetilde{A}_{u,v} \cdot R_{v,w} = m_i \cdot R_{v,w}}_{\text{with } v \text{ given by (89)}} = \begin{cases} m_i & \text{if } \textsc{Pre}(w) = \textsc{Suf}(u); \\ -m_i & \text{if } \textsc{Pre}(w) = \textsc{Suf}(Z^+(u)); \\ 0 & \text{otherwise} \end{cases} \tag{90}$$
$$= (RA)_{u,w} \text{ for this case, by Eq. (84).}$$

Case (iii): $u = Z_i$*:* Setting $u = Z_i$ in (87), we only need to consider v of the form:

$$v = \zeta_{i+1} \ldots \zeta_{i+k-2}\, c \ \in \ V_{i+1} \qquad \text{with } c \in \Omega_{i+k-1}. \tag{91}$$

There are two subcases: $w \neq Z_{i+1}$ or $w = Z_{i+1}$.

- *Case (iii), subcase* $w \neq Z_{i+1}$*:* By (88), either $v = w$ or $Z^+(v) = w$. But $Z^+(v) \neq w$ (for v of form (91), $Z^+(v)$ is either Z_{i+1} or undefined). So $v = w$ and sum (81b) is:

$$(\widetilde{A}R)_{u,w} = \widetilde{A}_{u,w} \cdot \underbrace{R_{w,w}}_{=1 \text{ by (70)}} = \widetilde{A}_{u,w}\,. \tag{92}$$

  Evaluate $\widetilde{A}_{u,w}$ by (78); since $u = Z_i$, then $u_0 = \zeta_i$, and since $w \neq Z_{i+1}$, only the first and third cases of (78) apply. Thus, for $u = Z_i$ and $w \in V_{i+1} \setminus \{Z_{i+1}\}$:

$$(\widetilde{A}R)_{u,w} = \begin{cases} m_i & \text{if } \textsc{Pre}(w) = \textsc{Suf}(u); \\ 0 & \text{otherwise.} \end{cases} \tag{93}$$

- *Case (iii), subcase* $w = Z_{i+1}$*:* All v of form (91) yield nonzero contributions to (81b):
  - If $c = \zeta_{i+k-1}$, then $v = Z_{i+1} = w$, which gives $\widetilde{A}_{u,v} = d_i$ and $R_{v,w} = 1$.
  - The other $q_{i+k-1} - 1$ choices of c in Ω_{i+k-1} give $v \neq Z_{i+1}$ and $Z^+(v) = Z_{i+1} = w$. Then $\widetilde{A}_{u,v} = m_i$ and $R_{v,w} = -1$.
  - Thus, sum (81b) is:

$$\begin{aligned}(\widetilde{A}R)_{u,w} &= d_i \cdot 1 + (q_{i+k-1} - 1)(m_i)(-1) \\ &= d_i - \underbrace{m_i q_{i+k-1}}_{=d_i \text{ by (29)}} + m_i = d_i - d_i + m_i = m_i.\end{aligned} \tag{94}$$

  Also, $\textsc{Pre}(w) = \textsc{Suf}(u) = Z_{i+1,k-2}$. Thus, (93) also holds for $u = Z_i$ and $w = Z_{i+1}$. Then (85) and (93) agree on case (iii), so $(RA)_{u,w} = (\widetilde{A}R)_{u,w}$ in case (iii). $\square$

7.4 Characteristic polynomial of the Laplacian

We stated the characteristic polynomial of the Laplacian matrix, L, in Theorem 12; we prove it in this section. Let $(\vec{m}, \vec{q}, k)$ be compatible parameters with $k \geqslant 2$. Use the change of basis (76), $\widetilde{L} = RLR^{-1}$, to compute

$$f(x) = \det(xI - L) = \det(xI - \widetilde{L}) = \sum_{\sigma \in \textsc{Sym}(V)} \textsc{Sign}(\sigma) \prod_{v \in V} (xI - \widetilde{L})_{v,\sigma(v)} \tag{95}$$

where $\textsc{Sym}(V)$ is the symmetric group on V and $\textsc{Sign}(\sigma)$ is the permutation sign. We'll show that this sum has at most two nonzero terms: $\sigma =$ identity and $\sigma = (Z_0, \ldots, Z_{p-1})$.

Set $M = xI - \widetilde{L}$. Diagonal entries of M are nonzero (as polynomials in x), while each off-diagonal entry of M is nonzero iff the corresponding entry of $\widetilde{L}$ is nonzero (since $M_{u,v} = -\widetilde{L}_{u,v}$ when $u \neq v$).

Lemma 24. *Suppose permutation $\sigma \in \textsc{Sym}(V)$ has $\prod_{u \in V} M_{u,\sigma(u)} \neq 0$, with $\sigma \neq$ identity. For $p \geqslant 2$, the only solution is $\sigma = (Z_0, \ldots, Z_{p-1})$. For $p = 1$, there is no such σ.*

Note: Permutation $\sigma = (Z_0, \ldots, Z_{p-1})$ is highlighted in the examples of $\widetilde{L}$ in Appendix A.

Proof. Since $\sigma \neq$ identity, it has a cycle $\gamma = (T_0, \ldots, T_{r-1})$ of length $r \geqslant 2$. Then $T_0, \ldots, T_{r-1}$ are distinct; $\sigma(T_i) = T_{i+1}$ for $0 \leqslant i \leqslant r-2$; and $\sigma(T_{r-1}) = T_0$.

The entries $M_{T_i,\sigma(T_i)}$ are off-diagonal, and are nonzero since their product is nonzero.

Let $e = \textsc{Phase}(T_0)$. By (79)–(80), nonzero off-diagonal entries $M_{T_i,T_{i+1}}$ require T_i and T_{i+1} to have consecutive phases. Thus, $\textsc{Phase}(T_i) = (e+i) \bmod p$ for $0 \leqslant i \leqslant r-1$.

Similarly, nonzero off-diagonal entry M_{T_{r-1},T_0} requires T_{r-1} and T_0 to have consecutive phases, so $e + (r-1) + 1 \equiv e \pmod{p}$, so $r \equiv 0 \pmod{p}$, so $p|r$.

For convenience, set $T_r = T_0$, so that $\sigma(T_i) = T_{i+1}$ for $0 \leqslant i \leqslant r-1$. Note $\textsc{Phase}(T_i) = (e+i) \pmod{p}$ holds for $i = r$ as well, since $e + r \equiv e + 0 \pmod{p}$.

Computing T_i's: Write $T_i = t_{i,0}\, t_{i,1} \ldots t_{i,k-2}$ (string of length $k-1$, with $t_{i,j} \in \Omega_{e+i+j}$).

In (79)–(80), nonzero off-diagonal entries require $t_{i,0} = \zeta_{e+i}$ and $\textsc{Suf}(T_i) = \textsc{Pre}(T_{i+1})$ for $0 \leqslant i \leqslant r-1$, so $t_{i,j} = t_{i+1,j-1}$ for $1 \leqslant j \leqslant k-2$.

Then $t_{i,1} = t_{i+1,0} = \zeta_{e+i+1}$ for $i \leqslant r-1$. (For $i = r-1$, that's $t_{r-1,1} = t_{r,0} = t_{0,0} = \zeta_e = \zeta_{e+r}$, since $p|r$ and subscripts on ζ are reduced mod p.)

Iterate to add one character at a time to every T_i, yielding $t_{i,j} = \zeta_{e+i+j}$, so $T_i = Z_{e+i}$.

Computing r and γ: If $r \geqslant p+1$, then $T_p = Z_e = T_0$; but elements of a permutation cycle can't repeat, so $r \leqslant p$. Combining $0 < r \leqslant p$ and $p|r$ gives $r = p$, so $\gamma = (Z_e, Z_{e+1}, \ldots, Z_{e+p-1}) = (Z_0, \ldots, Z_{p-1})$ (by rotating the cycle).

Computing σ: For $p \geqslant 2$, as there's only one nontrivial cycle, all other elements of V are in cycles of length 1, giving $\sigma = (Z_0, \ldots, Z_{p-1})$. But if $p = 1$, then $\gamma = (Z_0)$ does not have length at least 2 as required, and $\sigma = (Z_0)$ is the identity. □

Now we prove Theorem 12 for the case $k \geqslant 2$:

Theorem 25. *For compatible parameters $(\vec{m}, \vec{q}, k)$ with $k \geqslant 2$, the characteristic polynomial of L is*

$$\det(xI - L) = \left(\prod_{i=0}^{p-1}(x - d_i)^{n_i-1}\right) \cdot \left(\prod_{i=0}^{p-1}(x - d_i) - \prod_{i=0}^{p-1}(-d_i)\right). \tag{96}$$

Proof. First we treat the case $p \geqslant 2$. Eq. (95) evaluates $\det(xI - L) = \det(xI - \widetilde{L})$ as a certain sum over permutations, and Lemma 24 reduces this sum to two terms: $\sigma =$ identity and $\sigma = (Z_0, \ldots, Z_{p-1})$. Term $\sigma =$ identity is as follows (in $M = xI - \widetilde{L}$, evaluate entries of $\widetilde{L}$ using (79)):

$$\textsc{Sign}(\text{identity}) \cdot \prod_{v \in V} M_{v,v} = 1 \cdot \prod_{v \in V}\big(x - \textsc{Outdeg}(v)\big) = \prod_{i=0}^{p-1}(x - d_i)^{n_i}\ . \tag{97}$$

For $\sigma = (Z_0, Z_1, \ldots, Z_{p-1})$, we have $\textsc{Sign}(\sigma) = (-1)^{p-1}$, and the term is

$$\begin{aligned}
&(-1)^{p-1} \cdot \left(\prod_{i=0}^{p-1} M_{Z_i, Z_{i+1}}\right) \cdot \left(\prod_{v \in V \setminus \{Z_0, \ldots, Z_{p-1}\}} M_{v,v}\right) \\
&\qquad = (-1)^{p-1}\left(\prod_{i=0}^{p-1} d_i\right) \cdot \left(\prod_{v \in V \setminus \{Z_0, \ldots, Z_{p-1}\}} \big(x - \textsc{Outdeg}(v)\big)\right) \\
&\qquad = -\left(\prod_{i=0}^{p-1}(-d_i)\right)\left(\prod_{i=0}^{p-1}(x - d_i)^{n_i-1}\right)\ .
\end{aligned} \tag{98}$$

Note that $\prod_{i=0}^{p-1} d_i$ has p factors and $(-1)^{p-1}$ has $p-1$ factors, yielding a negative sign on the last line. Adding (97) and (98) gives

$$\begin{aligned}
\det(xI - \widetilde{L}) &= \prod_{i=0}^{p-1}(x - d_i)^{n_i} - \left(\prod_{i=0}^{p-1}(-d_i)\right)\left(\prod_{i=0}^{p-1}(x - d_i)^{n_i-1}\right) \\
&= \left(\prod_{i=0}^{p-1}(x - d_i)^{n_i-1}\right) \cdot \left(\prod_{i=0}^{p-1}(x - d_i) - \prod_{i=0}^{p-1}(-d_i)\right)\ .
\end{aligned} \tag{99}$$

This proves (96) for $p \geqslant 2$. Next, for $p = 1$, the only permutation with a nonzero term in (95) is $\sigma =$ identity, so $\det(xI - \widetilde{L})$ is as follows (evaluate entries of $\widetilde{L}$ using (80)):

$$\begin{aligned}
\det(xI - \widetilde{L}) &= \textsc{Sign}(\text{identity}) \cdot \prod_{v \in V} M_{v,v} \\
&= 1 \cdot \left(\prod_{v \in V \setminus \{Z_0\}} \big(x - \textsc{Outdeg}(v)\big)\right) \cdot \underbrace{(x - 0)}_{v = Z_0} = (x - d_0)^{n_0 - 1} \cdot x.
\end{aligned}$$

This proves (96) for $p = 1$. ☐

Next, we give an interpretation of the rightmost factor in (99):

Theorem 26. *Let* $(\vec{m}, \vec{q}, k)$ *be compatible parameters with* $k \geqslant 2$*. Let* $S = (S_{ij}) = (\widetilde{L}_{Z_i,Z_j})$ *for* $0 \leqslant i, j \leqslant p-1$ *(a* $p \times p$ *submatrix of* $\widetilde{L}$*). For* $p = 1$ *we have* $S = [0]$*, and for* $p \geqslant 2$*,*

$$S = \begin{bmatrix} d_0 & -d_0 & & & & \\ & d_1 & -d_1 & & & \\ & & \ddots & \ddots & & \\ & & & & d_{p-2} & -d_{p-2} \\ -d_{p-1} & & & & & d_{p-1} \end{bmatrix} \tag{100}$$

(a $p \times p$ *matrix; blank entries are 0's). Then*

$$\det(xI - S) = \prod_{i=0}^{p-1}(x - d_i) - \prod_{i=0}^{p-1}(-d_i). \tag{101}$$

Proof. Matrix S is the submatrix of $\widetilde{L} = D - \widetilde{A}$ with rows and columns indexed by $Z_0, \ldots, Z_{p-1}$. Compute those entries using (30) and (78) to get (100).

For $p \geqslant 2$, expand $\det(xI - S)$ along its first column. There are two nonzero terms, each a coefficient times the determinant of a triangular matrix, giving (101).

For $p = 1$, we have $S = [0]$ and $\det(xI - S) = \det[x] = x$, which agrees with (101). □

Theorem 27. *Eq. (96) also holds for compatible parameters* $(\vec{m}, \vec{q}, k)$ *with* $k = 1$*.*

Proof. Theorem 25 proved (96) for $k \geqslant 2$. Here, we'll evaluate both sides of (96) when $k = 1$ and show that they agree, but we'll need to use a different method than before.

For $k = 1$, Eq. (29) gives $d_i = m_{i-1}q_{i-1} = m_i q_i = d_{i+1}$ for all i, so $d_0 = \cdots = d_{p-1}$.

Every phase has just one vertex: $V_i = \{\emptyset_i\}$ and $n_i = 1$. There are m_i edges $\emptyset_i \xrightarrow{c} \emptyset_{i+1}$ for each $c \in \Omega_i$; thus, there are $m_i q_i = d_{i+1} = d_0$ edges from $\emptyset_i$ to $\emptyset_{i+1}$. For $p \geqslant 2$, we get these $p \times p$ matrices (blank entries are 0):

$$A = \begin{bmatrix} 0 & d_0 & & & \\ & 0 & d_0 & & \\ & & \ddots & \ddots & \\ & & & 0 & d_0 \\ d_0 & & & & 0 \end{bmatrix} \quad D = \begin{bmatrix} d_0 & & & & \\ & d_0 & & & \\ & & \ddots & & \\ & & & d_0 & \\ & & & & d_0 \end{bmatrix} \quad L = \begin{bmatrix} d_0 & -d_0 & & & \\ & d_0 & -d_0 & & \\ & & \ddots & \ddots & \\ & & & d_0 & -d_0 \\ -d_0 & & & & d_0 \end{bmatrix} \tag{102}$$

while for $p = 1$, we have $A = D = [d_0]$ and $L = D - A = [0]$. Thus, for $p \geqslant 1$, matrix L is a special case of matrix S of Theorem 26, with all d_i's equal to d_0.

Thus, by (101), $\det(xI - L) = (x - d_0)^p - (-d_0)^p$. This agrees with (96) (all factors $(x - d_i)^{n_i - 1}$ are $(x - d_i)^{1-1} = 1$, and the remaining factors give $(x - d_0)^p - (-d_0)^p$). □

Table 4: Parameters for reversed sequences in terms of parameters for forwards sequences. Renumber phases by subtracting them any $h \in \mathbb{P}$ (the canonical choice is $h = p - 1$).

Input parameters	**Parameters for reversed sequences**
# phases	$p^{\text{rev}} = p$
k-mer multiplicity, by phase	$m_i^{\text{rev}} = m_{h-k-i+1}$
As a vector:	$\vec{m}^{\text{rev}} = (m_0^{\text{rev}}, \dots, m_{p-1}^{\text{rev}})$
	$= (m_{h-k+1}, m_{h-k}, \dots, m_{h-k-p+2})$
Alphabet, by phase	$\Omega_i^{\text{rev}} = \Omega_{h-i}$
Alphabet size, by phase	$q_i^{\text{rev}} = q_{h-i}$
As a vector:	$\vec{q}^{\text{rev}} = (q_0^{\text{rev}}, \dots, q_{p-1}^{\text{rev}}) = (q_h, q_{h-1}, \dots, q_{h-p+1})$
Word size	$k^{\text{rev}} = k$
Derived parameters	
Vertex indegree, by phase	$d_i^{\text{rev}} = d_{h-k-i+2}$
# vertices per phase	$n_i^{\text{rev}} = n_{h-k-i+2}$
# k-mers per phase	$\theta_i^{\text{rev}} = \theta_{h-k-i+1}$

8 Reverse sequences

The *reverse string* of $s = s_0 \dots s_{r-1}$ is $s^{\text{rev}} = s_{r-1} \dots s_0$. The reverse of cyclic sequence (s) is $(s)^{\text{rev}} = (s^{\text{rev}})$. Reversing a phased multi de Bruijn sequence gives a phased multi de Bruijn sequence with new parameters (Table 4).

Choose any $h \in \mathbb{P}$ (the canonical choice is $h = p - 1$), and renumber the alphabets in the opposite order by subtracting from h: $\Omega_i^{\text{rev}} = \Omega_{h-i}$ and $q_i^{\text{rev}} = q_{h-i}$. Consider $y = y_0 \dots y_{k-1} \in K_i$. Then $y_{k-1} \in \Omega_{i+k-1} = \Omega_{h-k-i+1}^{\text{rev}}$, so $y^{\text{rev}} = y_{k-1} \dots y_0$ has reverse phase $h - k - i + 1$, giving $m_i^{\text{rev}} = m_{h-k-i+1}$ and $\theta_i^{\text{rev}} = \theta_{h-k-i+1}$. For $(k-1)$-mers, we similarly derive $n_i^{\text{rev}} = n_{h-k-i+2}$.

Let $G = G(\vec{m}, \vec{q}, \vec{k})$ and $G^{\text{rev}} = G(\vec{m}^{\text{rev}}, \vec{q}^{\text{rev}}, k)$ (with matrices A^{rev}, D^{rev}, L^{rev}, $\widetilde{A}^{\text{rev}}$, $\widetilde{L}^{\text{rev}}$). Then

$$A_{v,w} = (A^{\text{rev}})_{w^{\text{rev}}, v^{\text{rev}}} \qquad \text{(where } v, w \in V\text{)}. \tag{103}$$

Thus, A and A^{rev} are related by transposing the matrix and simultaneously permuting the rows and columns by reversing vertex strings. Matrices D and D^{rev}, and matrices L and L^{rev}, are also related by the operation in (103).

Matrices $\widetilde{L}$ and $\widetilde{L}^{\text{rev}}$ have at most one nonzero off-diagonal entry per column. The matrix $\widehat{L}$ with entries $\widehat{L}_{v,w} = (\widetilde{L}^{\text{rev}})_{w^{\text{rev}}, v^{\text{rev}}}$ (for $v, w \in V$) gives the Laplacian of G in a basis with at most one nonzero off-diagonal entry per row instead.

Edges going out from v in G correspond to the edges into v^{rev} in G^{rev}, and vice-versa, giving $d_i^{\text{rev}} = d_{h-k-i+2}$ by Lemma 10.

A Detailed examples of parameters and matrices

We give examples of the various notations and matrices in this paper. Appendix A.1 illustrates one phase (the classical case), and Appendix A.2 illustrates four phases, showing the block matrix structure not evident in the one phase case. In the matrices, zero entries are abbreviated as dots.

A.1 One phase: classical de Bruijn sequences and multi de Bruijn sequences

One phase ($p = 1$) gives multi de Bruijn sequences [18], a cyclic string over an alphabet Ω of size q which has all k-mers exactly m times; classical de Bruijn sequences are the subcase $m = 1$. Below we give parameters and matrices for $m \in \mathbb{Z}^+$, $q = 2$ (alphabet $\Omega = \{0, 1\}$), and $k = 3$. See Table 2 in the main text for formulas of derived parameters for multi de Bruijn sequences with arbitrary m, q, k.

Input parameters	**Classical de Bruijn**	**Multi de Bruijn**
# phases	$p = 1$	$p = 1$
k-mer multiplicity per phase	$\vec{m} = (1)$	$\vec{m} = (m)$
Alphabet size per phase	$\vec{q} = (2)$	$\vec{q} = (2)$
Word size	$k = 3$	$k = 3$
Derived parameters		
Vertex degrees per phase	$\vec{d} = (2)$	$\vec{d} = (2m)$
# vertices per phase	$\vec{n} = (4)$	$\vec{n} = (4)$
# k-mers per phase	$\vec{\theta} = (8)$	$\vec{\theta} = (8)$
Sequence length	$\ell = 8$	$\ell = 8m$

$D =$

	00	**01**	**10**	**11**
00	$2m$	$\cdot$	$\cdot$	$\cdot$
01	$\cdot$	$2m$	$\cdot$	$\cdot$
10	$\cdot$	$\cdot$	$2m$	$\cdot$
11	$\cdot$	$\cdot$	$\cdot$	$2m$

$A =$

	00	**01**	**10**	**11**
00	m	m	$\cdot$	$\cdot$
01	$\cdot$	$\cdot$	m	m
10	m	m	$\cdot$	$\cdot$
11	$\cdot$	$\cdot$	m	m

$\widetilde{A} = RAR^{-1} =$

	00	**01**	**10**	**11**
00	$\cdot$	$\cdot$	$\cdot$	$\cdot$
01	$\cdot$	$\cdot$	$\cdot$	$\cdot$
10	m	m	$\cdot$	$\cdot$
11	$\cdot$	$\cdot$	m	$2m$

$L =$

	00	**01**	**10**	**11**
00	m	$-m$	$\cdot$	$\cdot$
01	$\cdot$	$2m$	$-m$	$-m$
10	$-m$	$-m$	$2m$	$\cdot$
11	$\cdot$	$\cdot$	$-m$	m

$\widetilde{L} = RLR^{-1} =$

	00	**01**	**10**	**11**
00	$\boxed{2m}$	$\cdot$	$\cdot$	$\cdot$
01	$\cdot$	$\boxed{2m}$	$\cdot$	$\cdot$
10	$-m$	$-m$	$\boxed{2m}$	$\cdot$
11	$\cdot$	$\cdot$	$-m$	$\boxed{\cdot}$

(entries $\widetilde{L}_{u,\sigma(u)}$ boxed for $\sigma = (Z_0) =$ identity)

$R =$

	00	**01**	**10**	**11**
00	1	$\cdot$	-1	$\cdot$
01	$\cdot$	1	$\cdot$	-1
10	$\cdot$	$\cdot$	1	-1
11	$\cdot$	$\cdot$	$\cdot$	1

$R^{-1} =$

	00	**01**	**10**	**11**
00	1	$\cdot$	1	1
01	$\cdot$	1	$\cdot$	1
10	$\cdot$	$\cdot$	1	1
11	$\cdot$	$\cdot$	$\cdot$	1

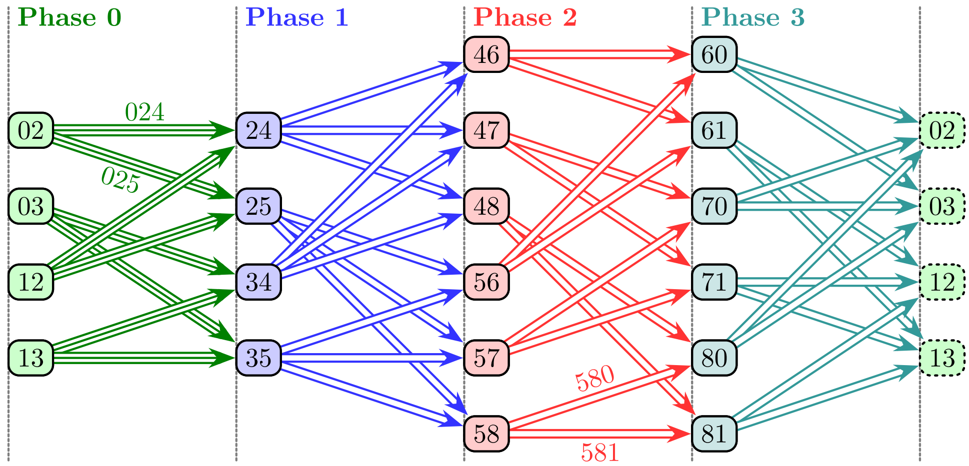

Figure 3: Phased Multi de Bruijn graph. Vertices and edges are colored by phase; e.g., vertex 02 and the three edges labelled 024 are in phase 0 (green). The four phase 0 vertices are drawn twice for legibility (solid green on left, dotted on right); to avoid repeating them, the graph could be redrawn in a circular layout. Each phase 0 edge has multiplicity $m_0 = 3$ (shown as triple lines), while edges in other phases have multiplicity $m_1 = m_2 = m_3 = 2$ (double lines). All edges are labelled by k-mers (3-mers), but for legibility, we only show a few edge labels as examples.

A.2 Four phase cyclic example

See Fig. 3 for the graph with the parameters below.

Input parameters	
# phases	$p = 4$
k-mer multiplicities	$\vec{m} = (3, 2, 2, 2)$
Alphabet sizes	$\vec{q} = (2, 2, 2, 3)$
Word size	$k = 3$

Derived parameters	
Total # vertices	$n = 20$
# vertices by phase	$\vec{n} = (4, 4, 6, 6)$
Vertex degrees by phase	$\vec{d} = (6, 6, 4, 4)$
# k-mers by phase	$\vec{\theta} = (8, 12, 12, 12)$
Sequence length	$\ell = \vec{m} \cdot \vec{\theta} = 96$
Compatibility	All $m_i \theta_i = 24$

Alphabet	**Alphabet size**	**Min**	**Max**
$\Omega_0 = \{0, 1\}$	$q_0 = \lvert\Omega_0\rvert = 2$	$\alpha_0 = 0$	$\zeta_0 = 1$
$\Omega_1 = \{2, 3\}$	$q_1 = \lvert\Omega_1\rvert = 2$	$\alpha_1 = 2$	$\zeta_1 = 3$
$\Omega_2 = \{4, 5\}$	$q_2 = \lvert\Omega_2\rvert = 2$	$\alpha_2 = 4$	$\zeta_2 = 5$
$\Omega_3 = \{6, 7, 8\}$	$q_3 = \lvert\Omega_3\rvert = 3$	$\alpha_3 = 6$	$\zeta_3 = 8$

$\boldsymbol{K_i}$ **=** $\boldsymbol{k}$**-mers by phase**	$\boldsymbol{\theta_i = \lvert K_i\rvert}$	**Multiplicity**
$K_0 = \{024, 025, 034, 035, 124, 125, 134, 135\}$	$\theta_0 = 8$	$m_0 = 3$
$K_1 = \{246, 247, 248, 256, 257, 258, 346, 347, 348, 356, 357, 358\}$	$\theta_1 = 12$	$m_1 = 2$
$K_2 = \{460, 461, 470, 471, 480, 481, 560, 561, 570, 571, 580, 581\}$	$\theta_2 = 12$	$m_2 = 2$
$K_3 = \{602, 603, 612, 613, 702, 703, 712, 713, 802, 803, 812, 813\}$	$\theta_3 = 12$	$m_3 = 2$

V_i = vertices by phase	$n_i = \lvert V_i \rvert$	Degree	Min vertex	Max vertex
$V_0 = \{02, 03, 12, 13\}$	$n_0 = 4$	$d_0 = 6$	$O_0 = 02$	$Z_0 = 13$
$V_1 = \{24, 25, 34, 35\}$	$n_1 = 4$	$d_1 = 6$	$O_1 = 24$	$Z_1 = 35$
$V_2 = \{46, 47, 48, 56, 57, 58\}$	$n_2 = 6$	$d_2 = 4$	$O_2 = 46$	$Z_2 = 58$
$V_3 = \{60, 61, 70, 71, 80, 81\}$	$n_3 = 6$	$d_3 = 4$	$O_3 = 60$	$Z_3 = 81$

Example sequence. Each sequence in $\mathcal{C}((3,2,2,2),(2,2,2,3),3)$ has length $\ell = 96$, comprised of $\ell/p = 96/4 = 24$ groups of $p = 4$ digits. Each group goes through all four phases consecutively. Spaces are shown for readability, but are not part of the sequence:

Positions $0 \ldots 47$: (0246 0258 0248 0258 1247 0356 0348 0357 0348 1358 0347 1256
Positions $48 \ldots 95$: 0356 1246 1358 1257 1346 0256 1257 1248 1346 1357 0247 1347)

Characteristic polynomial and eigenvalues of the Laplacian. By Theorems 13 and 25, the eigenvalues of the Laplacian matrix are as follows: Phases 0–3 give eigenvalues $\vec{d} = (6,6,4,4)$ with multiplicities $\vec{n} - \vec{1} = (3,3,5,5)$; Eq. (100) gives an eigenvalue 0; and the $p-1 = 3$ nonzero eigenvalues of S have a product 480, shown below. The total multiplicity of eigenvalue 6 is $3+3=6$ and the total multiplicity of 4 is $5+5=10$.

By Eq. (33), the characteristic polynomial of L is

$$\begin{aligned}\det(xI - L) &= \big((x-6)^3(x-6)^3(x-4)^5(x-4)^5\big) \\ &\quad \cdot \big((x-6)(x-6)(x-4)(x-4) - (-6)(-6)(-4)(-4)\big) \\ &= (x-6)^3(x-6)^3(x-4)^5(x-4)^5 \cdot \underbrace{\big(x^4 - 20x^3 + 148x^2 - 480x\big)}_{=\det(xI-S);\text{ see below}} . \qquad (104)\end{aligned}$$

The first four factors in this correspond to the four phases in the above table. The last factor can be computed as follows. Form matrix S by (100):

$$S = \begin{bmatrix} 6 & -6 & 0 & 0 \\ 0 & 6 & -6 & 0 \\ 0 & 0 & 4 & -4 \\ -4 & 0 & 0 & 4 \end{bmatrix}$$

Eigenvalues: $0,\ 10,\ 5 - i\sqrt{23},\ 5 + i\sqrt{23}$

Characteristic polynomial:
$\det(xI - S) = x^4 - 20x^3 + 148x^2 - 480x.$

The three nonzero eigenvalues of S have product 480. Two ways to compute this product without explicitly finding the eigenvalues are as the coefficient of x in $(-1)^{p-1}\det(xI - S)$, or by Eq. (34): $6 \cdot 6 \cdot 4 \cdot 4 \cdot \left(\frac{1}{6} + \frac{1}{6} + \frac{1}{4} + \frac{1}{4}\right) = 480$.

Matrices. All matrices are $n \times n = 20 \times 20$, indexed by vertices in lexicographical order.

Degree matrix (Eq. (30)):

$$D = \text{diagonal}(\underbrace{6,6,6,6,}_{\textbf{Phase 0}}\ \underbrace{6,6,6,6,}_{\textbf{Phase 1}}\ \underbrace{4,4,4,4,4,4,}_{\textbf{Phase 2}}\ \underbrace{4,4,4,4,4,4}_{\textbf{Phase 3}})$$

Adjacency matrix, A, and change of basis, $\widetilde{A} = RAR^{-1}$ (Eqs. (24) and (78)):

$A =$

	Phase 0				Phase 1				Phase 2						Phase 3					
	02	**03**	**12**	**13**	**24**	**25**	**34**	**35**	**46**	**47**	**48**	**56**	**57**	**58**	**60**	**61**	**70**	**71**	**80**	**81**
02	·	·	·	·	3	3	·	·	·	·	·	·	·	·	·	·	·	·	·	·
03	·	·	·	·	·	·	3	3	·	·	·	·	·	·	·	·	·	·	·	·
12	·	·	·	·	3	3	·	·	·	·	·	·	·	·	·	·	·	·	·	·
13	·	·	·	·	·	·	3	3	·	·	·	·	·	·	·	·	·	·	·	·
24	·	·	·	·	·	·	·	·	2	2	2	·	·	·	·	·	·	·	·	·
25	·	·	·	·	·	·	·	·	·	·	·	2	2	2	·	·	·	·	·	·
34	·	·	·	·	·	·	·	·	2	2	2	·	·	·	·	·	·	·	·	·
35	·	·	·	·	·	·	·	·	·	·	·	2	2	2	·	·	·	·	·	·
46	·	·	·	·	·	·	·	·	·	·	·	·	·	·	2	2	·	·	·	·
47	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	2	2	·	·
48	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	2	2
56	·	·	·	·	·	·	·	·	·	·	·	·	·	·	2	2	·	·	·	·
57	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	2	2	·	·
58	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	2	2
60	2	2	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
61	·	·	2	2	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
70	2	2	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
71	·	·	2	2	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
80	2	2	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
81	·	·	2	2	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·

$\widetilde{A} =$

	Phase 0				Phase 1				Phase 2						Phase 3					
	02	**03**	**12**	**13**	**24**	**25**	**34**	**35**	**46**	**47**	**48**	**56**	**57**	**58**	**60**	**61**	**70**	**71**	**80**	**81**
02	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
03	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
12	·	·	·	·	3	3	·	·	·	·	·	·	·	·	·	·	·	·	·	·
13	·	·	·	·	·	·	3	6	·	·	·	·	·	·	·	·	·	·	·	·
24	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
25	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
34	·	·	·	·	·	·	·	·	2	2	2	·	·	·	·	·	·	·	·	·
35	·	·	·	·	·	·	·	·	·	·	·	2	2	6	·	·	·	·	·	·
46	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
47	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
48	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
56	·	·	·	·	·	·	·	·	·	·	·	·	·	·	2	2	·	·	·	·
57	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	2	2	·	·
58	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	2	4
60	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
61	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
70	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
71	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
80	2	2	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
81	·	·	2	4	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·

Laplacian matrix, $\boldsymbol{L}$, and change of basis, $\widetilde{\boldsymbol{L}} = \boldsymbol{RLR^{-1}}$ (Eqs. (31) and (79)):

$L =$

	Phase 0				**Phase 1**				**Phase 2**						**Phase 3**					
	02	**03**	**12**	**13**	**24**	**25**	**34**	**35**	**46**	**47**	**48**	**56**	**57**	**58**	**60**	**61**	**70**	**71**	**80**	**81**
02	6	·	·	·	−3	−3	·	·	·	·	·	·	·	·	·	·	·	·	·	·
03	·	6	·	·	·	·	−3	−3	·	·	·	·	·	·	·	·	·	·	·	·
12	·	·	6	·	−3	−3	·	·	·	·	·	·	·	·	·	·	·	·	·	·
13	·	·	·	6	·	·	−3	−3	·	·	·	·	·	·	·	·	·	·	·	·
24	·	·	·	·	6	·	·	·	−2	−2	−2	·	·	·	·	·	·	·	·	·
25	·	·	·	·	·	6	·	·	·	·	·	−2	−2	−2	·	·	·	·	·	·
34	·	·	·	·	·	·	6	·	−2	−2	−2	·	·	·	·	·	·	·	·	·
35	·	·	·	·	·	·	·	6	·	·	·	−2	−2	−2	·	·	·	·	·	·
46	·	·	·	·	·	·	·	·	4	·	·	·	·	·	−2	−2	·	·	·	·
47	·	·	·	·	·	·	·	·	·	4	·	·	·	·	·	·	−2	−2	·	·
48	·	·	·	·	·	·	·	·	·	·	4	·	·	·	·	·	·	·	−2	−2
56	·	·	·	·	·	·	·	·	·	·	·	4	·	·	−2	−2	·	·	·	·
57	·	·	·	·	·	·	·	·	·	·	·	·	4	·	·	·	−2	−2	·	·
58	·	·	·	·	·	·	·	·	·	·	·	·	·	4	·	·	·	·	−2	−2
60	−2	−2	·	·	·	·	·	·	·	·	·	·	·	·	4	·	·	·	·	·
61	·	·	−2	−2	·	·	·	·	·	·	·	·	·	·	·	4	·	·	·	·
70	−2	−2	·	·	·	·	·	·	·	·	·	·	·	·	·	·	4	·	·	·
71	·	·	−2	−2	·	·	·	·	·	·	·	·	·	·	·	·	·	4	·	·
80	−2	−2	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	4	·
81	·	·	−2	−2	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	4

$\widetilde{L} =$

02	$\boxed{6}$	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
03	·	$\boxed{6}$	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
12	·	·	$\boxed{6}$	·	−3	−3	·	·	·	·	·	·	·	·	·	·	·	·	·	·
13	·	·	·	6	·	·	−3	$\boxed{-6}$	·	·	·	·	·	·	·	·	·	·	·	·
24	·	·	·	·	$\boxed{6}$	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
25	·	·	·	·	·	$\boxed{6}$	·	·	·	·	·	·	·	·	·	·	·	·	·	·
34	·	·	·	·	·	·	$\boxed{6}$	·	−2	−2	−2	·	·	·	·	·	·	·	·	·
35	·	·	·	·	·	·	·	6	·	·	·	−2	−2	$\boxed{-6}$	·	·	·	·	·	·
46	·	·	·	·	·	·	·	·	$\boxed{4}$	·	·	·	·	·	·	·	·	·	·	·
47	·	·	·	·	·	·	·	·	·	$\boxed{4}$	·	·	·	·	·	·	·	·	·	·
48	·	·	·	·	·	·	·	·	·	·	$\boxed{4}$	·	·	·	·	·	·	·	·	·
56	·	·	·	·	·	·	·	·	·	·	·	$\boxed{4}$	·	·	−2	−2	·	·	·	·
57	·	·	·	·	·	·	·	·	·	·	·	·	$\boxed{4}$	·	·	·	−2	−2	·	·
58	·	·	·	·	·	·	·	·	·	·	·	·	·	4	·	·	·	·	−2	$\boxed{-4}$
60	·	·	·	·	·	·	·	·	·	·	·	·	·	·	$\boxed{4}$	·	·	·	·	·
61	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	$\boxed{4}$	·	·	·	·
70	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	$\boxed{4}$	·	·	·
71	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	$\boxed{4}$	·	·
80	−2	−2	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	$\boxed{4}$	·
81	·	·	−2	$\boxed{-4}$	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	4

Note: We boxed entries $\widetilde{L}_{u,\sigma(u)}$ for $\sigma = (Z_0, Z_1, Z_2, Z_3) = (13, 35, 58, 81)$ and all $u \in V$; note $\sigma(u) = u$ for vertices not of the form Z_i, e.g., $\sigma(02) = 02$. See Lemma 24.

Change of basis matrices R and R^{-1} (Eqs. (70) and (71)):

$R =$

	Phase 0				Phase 1				Phase 2						Phase 3					
	02	**03**	**12**	**13**	**24**	**25**	**34**	**35**	**46**	**47**	**48**	**56**	**57**	**58**	**60**	**61**	**70**	**71**	**80**	**81**
02	1	·	−1	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
03	·	1	·	−1	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
12	·	·	1	−1	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
13	·	·	·	1	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
24	·	·	·	·	1	·	−1	·	·	·	·	·	·	·	·	·	·	·	·	·
25	·	·	·	·	·	1	·	−1	·	·	·	·	·	·	·	·	·	·	·	·
34	·	·	·	·	·	·	1	−1	·	·	·	·	·	·	·	·	·	·	·	·
35	·	·	·	·	·	·	·	1	·	·	·	·	·	·	·	·	·	·	·	·
46	·	·	·	·	·	·	·	·	1	·	·	−1	·	·	·	·	·	·	·	·
47	·	·	·	·	·	·	·	·	·	1	·	·	−1	·	·	·	·	·	·	·
48	·	·	·	·	·	·	·	·	·	·	1	·	·	−1	·	·	·	·	·	·
56	·	·	·	·	·	·	·	·	·	·	·	1	·	−1	·	·	·	·	·	·
57	·	·	·	·	·	·	·	·	·	·	·	·	1	−1	·	·	·	·	·	·
58	·	·	·	·	·	·	·	·	·	·	·	·	·	1	·	·	·	·	·	·
60	·	·	·	·	·	·	·	·	·	·	·	·	·	·	1	·	·	·	−1	·
61	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	1	·	·	·	−1
70	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	1	·	−1	·
71	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	1	·	−1
80	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	1	−1
81	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	1

$R^{-1} =$

	Phase 0				Phase 1				Phase 2						Phase 3					
	02	**03**	**12**	**13**	**24**	**25**	**34**	**35**	**46**	**47**	**48**	**56**	**57**	**58**	**60**	**61**	**70**	**71**	**80**	**81**
02	1	·	1	1	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
03	·	1	·	1	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
12	·	·	1	1	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
13	·	·	·	1	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
24	·	·	·	·	1	·	1	1	·	·	·	·	·	·	·	·	·	·	·	·
25	·	·	·	·	·	1	·	1	·	·	·	·	·	·	·	·	·	·	·	·
34	·	·	·	·	·	·	1	1	·	·	·	·	·	·	·	·	·	·	·	·
35	·	·	·	·	·	·	·	1	·	·	·	·	·	·	·	·	·	·	·	·
46	·	·	·	·	·	·	·	·	1	·	·	1	·	1	·	·	·	·	·	·
47	·	·	·	·	·	·	·	·	·	1	·	·	1	1	·	·	·	·	·	·
48	·	·	·	·	·	·	·	·	·	·	1	·	·	1	·	·	·	·	·	·
56	·	·	·	·	·	·	·	·	·	·	·	1	·	1	·	·	·	·	·	·
57	·	·	·	·	·	·	·	·	·	·	·	·	1	1	·	·	·	·	·	·
58	·	·	·	·	·	·	·	·	·	·	·	·	·	1	·	·	·	·	·	·
60	·	·	·	·	·	·	·	·	·	·	·	·	·	·	1	·	·	·	1	1
61	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	1	·	·	·	1
70	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	1	·	1	1
71	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	1	·	1
80	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	1	1
81	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	1

Constructing matrices R and R^{-1}: These are block diagonal matrices partitioned by vertex phases. When u and v have different phases (off-diagonal blocks), $R_{u,v} = (R^{-1})_{u,v} = 0$. Below, we construct the on-diagonal blocks (equal phases).

Matrix R: We build this matrix row-by-row. Let $u \in V_i$. Per Eq. (70), row u of R has either one or two nonzero entries:

- If $u = Z_i$, all entries in row u are 0, except $R_{u,u} = 1$. Note that $Z^+(u)$ is undefined.
- If $u \neq Z_i$, then $R_{u,u} = 1$; $R_{u,x} = -1$ for $x = Z^+(u)$; and the rest of row u is 0.

Recall that $Z_0 = 13$, $Z_1 = 35$, $Z_2 = 58$, and $Z_3 = 81$ (as strings). In the table below, for each $u \in V_i$, we first compute $h = \text{ZLEN}(u)$, the length of largest prefix of u that matches the corresponding positions in Z_i. Then we compute $x = Z^+(u)$ by changing position h of u to ζ_{i+h}. This position is underlined in x in the table below. But when $u = Z_i$, we have $h = k-1$ (here, $h = 2$); that's beyond the end of u, so $Z^+(u)$ is undefined (denoted "–" in this table).

	Phase 0				Phase 1				Phase 2						Phase 3					
u	02	03	12	13	24	25	34	35	46	47	48	56	57	58	60	61	70	71	80	81
$h = \text{ZLEN}(u)$	0	0	1	2	0	0	1	2	0	0	0	1	1	2	0	0	0	0	1	2
$x = Z^+(u)$	$\underline{1}2$	$\underline{1}3$	$1\underline{3}$	–	$\underline{3}4$	$\underline{3}5$	$3\underline{5}$	–	$\underline{5}6$	$\underline{5}7$	$\underline{5}8$	$5\underline{8}$	$5\underline{8}$	–	$\underline{8}0$	$\underline{8}1$	$\underline{8}0$	$\underline{8}1$	$8\underline{1}$	–

For example, row $u = 47$ has nonzero entries $R_{47,47} = 1$ and $R_{47,57} = -1$, and all other entries are zero. Row $u = Z_2 = 58$ only has one nonzero entry, $R_{58,58} = 1$.

Also note that within each phase, $\text{ZLEN}(u)$ is weakly increasing, by Lemma 18(d).

Matrix R^{-1}: We build this matrix column-by-column. Let $v \in V_i$ and set $h = \text{ZLEN}(v)$. Per Eq. (71), the nonzero entries in column v are $(R^{-1})_{u,v} = 1$, where u satisfies:

- $u \in V_i$
- The first h positions of u are arbitrary in their phases (denoted '?' in the table below).
- The remaining $k - 1 - h$ positions of u and v agree: $\text{SUF}_{k-1-h}(u) = \text{SUF}_{k-1-h}(v)$.

Each $u \in V$ satisfying these has $(R^{-1})_{u,v} = 1$; all other $u \in V$ have $(R^{-1})_{u,v} = 0$.

For example, $v = 57 \in V_2$ gives the pattern $u =$ "?7" for $u \in V_2$. So column $v = 57$ has $(R^{-1})_{u,v} = 1$ for $u \in \{47, 57\}$ while $(R^{-1})_{u,v} = 0$ for all other $u \in V$.

For $v = 58 \in V_2$, we have pattern $u =$ "??" with $u \in V_2$. So column $v = 58$ has $(R^{-1})_{u,v} = 1$ for all $u \in V_2$, while $(R^{-1})_{u,v} = 0$ for all other $u \in V$.

	Phase 0				Phase 1				Phase 2						Phase 3					
v	02	03	12	13	24	25	34	35	46	47	48	56	57	58	60	61	70	71	80	81
$h = \text{ZLEN}(v)$	0	0	1	2	0	0	1	2	0	0	0	1	1	2	0	0	0	0	1	2
u pattern	02	03	?2	??	24	25	?4	??	46	47	48	?6	?7	??	60	61	70	71	?0	??

References

[1] F. Bonhoure, Y. Dallery, and W.J. Stewart. Algorithms for periodic Markov chains. In *Linear algebra, Markov chains, and queueing models*, pages 71–88. New York: Springer-Verlag, 1993.

[2] F. Chung, P. Diaconis, and R. Graham. Universal cycles for combinatorial structures. *Discrete Math.*, 110(1-3):43–59, 1992.

[3] E. Cinlar. Periodicity in Markov renewal theory. *Adv. Appl. Probab.*, 6:61–78, 1974.

[4] R. Dawson and I.J. Good. Exact Markov probabilities from oriented linear graphs. *Ann. Math. Stat.*, 28:946–956, 1957.

[5] N.G. de Bruijn. A combinatorial problem. *Proc. Akad. Wet. Amsterdam*, 49:758–764, 1946.

[6] N.G. de Bruijn. 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. Technological University Eindhoven, T.H.-Report 75-WSK-06, June 1975.

[7] A. de Rivière. Question 48. *l'Intermédiare des Mathématiciens*, 1:19–20, 1894.

[8] H. Fredricksen. A survey of full length nonlinear shift register cycle algorithms. *SIAM Rev.*, 24:195–221, 1982.

[9] N. Hartsfield and J.S. Werth. Spanning trees of the complete bipartite graph. In *Topics in combinatorics and graph theory (Oberwolfach, 1990)*, pages 339–346. Physica, Heidelberg, 1990.

[10] P.M. Higgins. Burrows-Wheeler transformations and de Bruijn words. *Theor. Comput. Sci.*, 457:128–136, 2012.

[11] G.H. Hurlbert. *Universal cycles: On beyond de Bruijn*. ProQuest LLC, Ann Arbor, MI, 1990. Ph.D. Thesis, Rutgers University.

[12] S. Kirkland. The group inverse associated with an irreducible periodic nonnegative matrix. *SIAM J. Matrix Anal. Appl.*, 16(4):1127–1134, 1995.

[13] R. Onodera. On the Number of Trees in a Complete N-Partite Graph. *Matrix Tensor Quart.*, 23(4):142–146, 1973.

[14] J. Pajarinen and J. Peltonen. Periodic Finite State Controllers for Efficient POMDP and DEC-POMDP Planning. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc., 2011.

[15] C.F. Sainte-Marie. Solution to question 48. *l'Intermédiare des Mathématiciens*, 1:107–110, 1894.

[16] H.I. Scoins. The number of trees with nodes of alternate parity. *Proc. Cambridge Philos. Soc.*, 58:12–16, 1962.

[17] R.P. Stanley. *Enumerative Combinatorics, Volume 2*. Cambridge University Press, Cambridge, 1999.

[18] G. Tesler. Multi de Bruijn sequences. *J. Comb.*, 8(3):439–474, 2017.

[19] W.T. Tutte. The dissection of equilateral triangles into equilateral triangles. *Proc. Camb. Philos. Soc.*, 44:463–482, 1948.

[20] T. van Aardenne-Ehrenfest and N.G. de Bruijn. Circuits and trees in oriented linear graphs. *Simon Stevin*, 28:203–217, 1951.