ARS MATHEMATICA CONTEMPORANEA

Volume 10, Number 1, Spring/Summer 2016, Pages 1–209

Covered by:
Mathematical Reviews
Zentralblatt MATH
COBISS
SCOPUS
Science Citation Index-Expanded (SCIE)
Web of Science
ISI Alerting Service
Current Contents/Physical, Chemical & Earth Sciences (CC/PC & ES)

The University of Primorska

The Society of Mathematicians, Physicists and Astronomers of Slovenia
The Institute of Mathematics, Physics and Mechanics

The publication is partially supported by the Slovenian Research Agency from the Call for co-financing of scientific periodical publications.

The ‘$2n - 1$ rule’

The ‘$2n-1$ rule’ is an old (and innocent) joke, often told by hard working honest mathematicians, about building a fence between good and bad mathematics. According to this rule, n-th rate mathematicians feel comfortable only when surrounded by $(2n-1)$-st rate mathematicians, as the latter do not pose a threat to the former. This rule has a fixed point only when $n = 1$, suggesting that first-rate mathematicians interact only with other first-rate mathematicians. An alternative version replaces $(2n-1)$-st rate mathematicians by k-th rate mathematicians for $k \geq 2n-1$. Clearly the $2n-1$ rule is a sociological rather than a mathematical rule, and is prone to exceptions.

One exception is that the advisor/supervisor of a first-rate mathematician need not be a first-rate mathematician — he or she may be an average mathematician, blessed by an exceptionally good student. Another exception, even more convincing, can be found in the history of mathematics. Many mathematicians can find big names such as Gauss or Newton among their academic predecessors (with help from the Mathematics Genealogy Project), and in particular, this shows that a first-rate mathematician can end up having an academic descendant who is not a first-rate mathematician himself/herself. And of course the $2n-1$ rule would then make it likely that at some value of n (much) greater than 1, some bad mathematics is bound to happen. A third exception comes in the form of co-authorship. If the $2n-1$ rule held without exceptions, then a lot of multi-author papers would be written by first-rate mathematicians only.

We are now seeing a revival of belief in the $2n-1$ rule, in the world of mathematical publishing. Most papers in n-th rate journals are cited in $(2n-1)$-th rate journals. This is supported by the aggressive growth of new mathematical journals, many of them published by predatory publishers, and feeding on the APC ‘Open Access’ publishing policy. It seems there are now too many mathematical journals, and too much mathematics is being published these days.

If taken *cum grano salis* (that is, with a grain of salt), the $2n-1$ rule is an amusing observation about quality in mathematics. Most mathematicians silently agree who are the first-rate mathematicians, and which journals are indeed first-rate journals. A problem can arise in isolated environments, however, when a few individuals declare themselves as first-rate mathematicians, and then by following the $2n-1$ rule, abuse their power. This is the main reason for not trusting blindly the $2n-1$ rule.
In our journal we do not believe in this kind of ranking. There is good mathematics and bad mathematics, and there are good journals and bad journals. And we are very happy that *Ars Mathematica Contemporanea* is a good journal, publishing good mathematics.

Dragan Marušič and Tomaž Pisanski
Editors In Chief

ARS MATHEMATICA
CONTEMPORANEA

Contents

Volume 10, Number 1, Spring/Summer 2016, Pages 1–209

Also available at http://amc-journal.eu
ISSN 1855-3966 (printed edn.), ISSN 1855-3974 (electronic edn.)
ARS MATHEMATICA CONTEMPORANEA 10 (2016) 1–8

Forbidden configurations: Finding the number predicted by the Anstee-Sali conjecture is NP-hard

Miguel Raggi *

Centro de Ciencias Matemáticas and Escuela Nacional de Estudios Superiores
Universidad Nacional Autónoma de México

Received 31 January 2013, accepted 23 June 2014, published online 24 January 2015

Abstract

Let F be a (possibly non-simple) hypergraph and let $\text{forb}(m, F)$ denote the maximum number of edges a simple hypergraph with m vertices can have if it doesn't contain F as a subhypergraph. A conjecture of Anstee and Sali predicts the asymptotic behaviour of $\text{forb}(m, F)$ for fixed F. In this paper we prove that even finding this predicted asymptotic behaviour is an NP-hard problem, meaning that if the Anstee-Sali conjecture were true, finding the asymptotics of $\text{forb}(m, F)$ would be NP-hard.

Keywords: Forbidden configuration, hypergraph, trace, NP-hard, NP-complete, Anstee-Sali conjecture.

Math. Subj. Class.: 05D05, 68R01

1 Introduction

This paper considers an extremal problem in hypergraph theory that results as a natural generalization of Turán's famous problem.

Some of the most celebrated extremal results are those of Erdős, Stone and Simonovits ([11], [10]). They consider the following problem: Given $m \in \mathbb{N}$ and a graph F, find the maximum number of edges in a graph on m vertices that avoids having a subgraph isomorphic to F.

There are a number of ways to generalize this to hypergraphs. A k-uniform hypergraph is one in which each edge has size k. Some view k-uniform hypergraphs as the most natural generalization of a graph (a graph is a 2-uniform hypergraph) and one might also generalize

*Research supported by DGAPA Postdoctoral scholarship, UNAM.
E-mail address: mraggi@gmail.com (Miguel Raggi)

the forbidden subgraph problem to a forbidden k-uniform subhypergraph problem. There are both asymptotic and exact results (e.g. [9], [13], [12]).

Forbidden Configurations is a different (but also natural) generalization that is studied mainly by Richard Anstee and his colaborators. We consider the following problem: Given $m \in \mathbb{N}$ and a hypergraph F, find the maximum number of edges in a *simple* hypergraph (*i.e.* no repeated edges) H on m vertices that avoids having a subhypergraph isomorphic to F. Surveys on the topic can be found in [1] and [14].

We find it convenient to use the language of matrices to describe hypergraphs: Each column of a $\{0,1\}$-matrix can be viewed as an incidence vector on the set of rows.

Definition 1.1. If α is a column and A a matrix, define $\lambda(\alpha, A)$ to be the multiplicity of α in A. Define a matrix to be **simple** if it is a $\{0,1\}$-matrix with no repeated columns (that is, if $\lambda(\alpha, A) \leq 1$ for every column α).

Note that an $m \times n$ simple matrix corresponds to a **simple hypergraph** (or **set system**) on m vertices with n distinct edges, where we allow the "empty edge".

Definition 1.2. When A is a $\{0,1\}$-matrix, we denote by $\|A\|$ the number of columns in A (which is the cardinality of the associated set system).

Definition 1.3. Let A and B be $\{0,1\}$-matrices with the same number of rows. Define the **concatenation** $[A|B]$ to be the configuration that results from taking all columns of A together with all columns of B. For $t \in \mathbb{N}$, we define the product

$$t \cdot A := [\underbrace{A \mid A \mid \cdots \mid A}_{t \text{ times}}].$$

Our objects of study are $\{0,1\}$-matrices with row and column order information stripped from them.

Definition 1.4. Two $\{0,1\}$-matrices are said to be equivalent if one is a row and column permutation of the other. An equivalence class is called a **configuration**.

Abusing notation, we will commonly use matrices and their corresponding configurations interchangeably.

Definition 1.5. For a configuration F and a $\{0,1\}$-matrix A (or a configuration A), we say that F is a **subconfiguration** of A, and write $F \prec A$ if there is a representative of F which is a submatrix of A. We say A **has no configuration** F (or **doesn't contain** F **as a configuration**) if F is not a subconfiguration of A. Let $\mathrm{Avoid}(m, F)$ denote the set of all simple matrices on m-rows with no subconfiguration F.

Our main extremal problem is to compute

$$\mathrm{forb}(m, F) = \max_{A}\{\|A\| \ : \ A \in \mathrm{Avoid}(m, F)\}.$$

Perhaps some examples are useful:

- $\mathrm{forb}\left(m, \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = m + 1$, since we can take all columns with at most one 1.

- $\text{forb}\left(m, \begin{bmatrix}1\\0\end{bmatrix}\right) = 2$, since we may only take the column of 1's and the column of 0's (*i.e.* the empty set and the complete set).
- $\text{forb}\left(m, \begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}\right) = \binom{m}{2} + \binom{m}{1} + \binom{m}{0}$, by taking all columns with at most two 1's. The proof that this is indeed the maximum is easy and can be found in [14].

Let A^c denote the $\{0,1\}$-complement of A (replace every 0 in A by a 1 and every 1 by a 0). Note that $\text{forb}(m, F) = \text{forb}(m, F^c)$.

Remark 1.6. Let F and G be configurations such that $F \prec G$. Then $\text{forb}(m, F) \leq \text{forb}(m, G)$.

We say a column α has **column sum** t if it has exactly t ones. Let $\mathbf{0}_m$ denote the column with m rows, all of them zeros. Similarly, let $\mathbf{1}_m$ denote the column of m ones.

For a set of rows S, we let $A|_S$ denote the submatrix of A given by restricting the rows of A to only those in S.

An important general result due to Füredi applies to simple and to non-simple configurations.

Theorem 1.7 (Z. Füredi)**.** *Let F be a given k-rowed $\{0,1\}-$matrix. Then* $\text{forb}(m, F)$ *is in $O(m^k)$.*

We desire more accurate asymptotic bounds. Anstee and Sali conjectured that the best asymptotic bounds can be achieved with certain *product constructions*.

Definition 1.8. Let A and B be $\{0,1\}$-matrices. We define the **product** $A \times B$ by taking each column of A and putting it on top of every column of B. Here is an example of a product:

$$A = \begin{bmatrix}0 & 1 & 1\\0 & 0 & 1\end{bmatrix},\ B = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix} \implies \begin{matrix}A\\ \times\\ B\end{matrix} = \left[\begin{array}{cc|cc|cc}0 & 0 & 1 & 1 & 1 & 1\\0 & 0 & 0 & 0 & 1 & 1\\ \hline 1 & 0 & 1 & 0 & 1 & 0\\0 & 1 & 0 & 1 & 0 & 1\end{array}\right].$$

Note that this is a well defined operation in configurations.

We are interested in asymptotic bounds for $\text{forb}(m, F)$. Let $\mathcal{I}_m$ be the $m \times m$ identity matrix, $\mathcal{I}_m^c$ be the $\{0,1\}$-complement of $\mathcal{I}_m$ (all ones except for the diagonal) and let $\mathcal{T}_m$ be the tower matrix: a matrix corresponding to a maximum chain in the partially ordered set of the power set of the vertices. For example,

$$\mathcal{T}_4 = \begin{bmatrix}0 & 1 & 1 & 1 & 1\\0 & 0 & 1 & 1 & 1\\0 & 0 & 0 & 1 & 1\\0 & 0 & 0 & 0 & 1\end{bmatrix}.$$

Anstee and Sali conjectured that the asymptotically "best" constructions avoiding a single configuration would be products of $\mathcal{I}, \mathcal{I}^c$ and $\mathcal{T}$.

Definition 1.9. Let F be a configuration. Let

$$P_r(a,b,c) := \underbrace{\mathcal{I}_r \times ... \times \mathcal{I}_r}_{a \text{ times}} \times \underbrace{\mathcal{I}_r^c \times ... \times \mathcal{I}_r^c}_{b \text{ times}} \times \underbrace{\mathcal{T}_r \times ... \times \mathcal{T}_r}_{c \text{ times}},$$

Define $X(F)$ to be the largest number such that there exist numbers $a, b, c \in \mathbb{N}$ with $a + b + c = X(F)$ such that for all $r \in \mathbb{N}$,

$$F \nprec P_r(a,b,c).$$

Conjecture 1.10. *[8] Let F be a configuration. Then* $\operatorname{forb}(m, F)$ *is in* $\Theta(m^{X(F)})$.

Observe that $X(F)$ is always an integer. Also note that $\|P_r(a,b,c)\| = r^{a+b} \cdot (r+1)^c \in \Theta(r^{X(F)})$, so by taking $r = \lceil m/X(F) \rceil$ (and perhaps deleting a constant number of rows and therefore columns in case $X(F) \nmid m$), we have that $\|P_r(a,b,c)\| \in \Omega(m^{X(F)})$. So the fact that $\operatorname{forb}(m,F) \in \Omega(m^{X(F)})$ is built into the conjecture.

Thus, in order to prove the conjecture, all that would be required would be to prove that $\operatorname{forb}(m,F) \in O(m^{X(F)})$ for every F. A disproof could be potentially easier, as only a counterexample would be required.

The conjecture has been proven for all $k \times \ell$ configurations F with $k \in \{1,2,3\}$ and many others cases in various papers. The proofs for $k = 2$ are in [4], for $k = 3$ in [4], [2], [8]. For $\ell = 2$, the conjecture was verified in [5]. For $k = 4$, all cases either when the conjecture predicts a cubic bound for F or when F is simple were completed in [3]. For $k = 4$ and F non-simple, there are three boundary cases with quadratic bounds, one of which is established in [6]. For $k \in \{5, 6\}$ some results can be found in [7].

Anstee has long conjectured that even finding $X(F)$ given F was not a trivial task, and the specific question of its NP-hardness was conjectured in [1], [14]. In this paper we settle this question: finding X is indeed NP-hard. We also note that one of the decision versions associated with this optimization problem is NP-complete, adding this function to the long list of functions known to be NP-complete, with the interesting plus that this function is conjectured to give the *exponent* of the asymptotic growth of forb.

For relatively small configurations F we have a computer program that yields the answer (relatively) quickly. The source code (in C++) can be freely downloaded from:

`http://matmor.unam.mx/~mraggi/`

This program can compute $X(F)$ for F having less than $\sim$10 rows in just a few minutes. This task takes merely exponential time, not doubly exponential (as it is often the case with forbidden configuration problems). This program was written to perform many tasks other than finding $X(F)$. A description of the algorithms used can be found in [14].

2 Results

There are two natural decision problems associated with $X(F)$: Given F and k as inputs,

1. Is it true that $X(F) < k$?
2. Is it true that $X(F) \geq k$?

We prove that the first of the two decision problems is in NP by exhibiting a certificate which can be checked in polynomial time.

The main result of this paper is the following:

Theorem 2.1. *Finding $X(F)$ is an NP-hard problem. In other words, should a polynomial-time algorithm exist for finding $X(F)$ given F, then every problem in NP could be solved in polynomial time. Furthermore, the problem "given F and k, is $X(F) < k$?" is in NP.*

Before proving this theorem, we need a few results.

Proposition 2.2. *Let F be a configuration with n rows. Then $X(F) \leq n$.*

Proof. Indeed, assume for the sake of contradiction that a, b, and c are such that $a+b+c = n+1$ and $F \not\prec P_r(a,b,c)$. Every configuration of $P_r(a,b,c)$ contains $[01]$, therefore any $\{0,1\}$-column can be formed with the first n configurations of $P_r(a,b,c)$. The $n+1$-th matrix ensures the columns of F with high multiplicity get repeated as many times as needed. □

This observation in particular implies that if a polynomial time algorithm existed for any of the two decision versions of the problem, then we'd have a polynomial time algorithm for finding $X(F)$, which together with Theorem 2.1 would make the decision version of finding $X(F)$ an NP-complete problem.

A simple (but surprising) corollary of Conjecture 1.10, if it were true, would be that repeating columns more than twice in F has no effect on the asymptotic behavior of $\mathrm{forb}(m,F)$. In other words, assuming the conjecture were true, the multiplicity of a column in a configuration would not affect the asymptotic bound and, asymptotically, it would only matter if a column is not there (has multiplicity 0), appears once (has multiplicity 1), or appears "multiple times" (has multiplicity 2 or more). Formally,

Proposition 2.3. *Let $F_t = [G|t \cdot H]$ with G and H simple $\{0,1\}$-matrices that have no columns in common. Then $X(F_2) = X(F_t)$ for all $t \geq 2$. In particular, if the conjecture were true, then $\mathrm{forb}(m,F_t)$ and $\mathrm{forb}(m,F_2)$ would have the same asymptotic behavior.*

Proof. It suffices to show that given t, G, H, a, b and c, there exists an R such that for every $r \geq R$, we have

$$F_2 = [G|2 \cdot H] \prec P_r(a,b,c) \iff F_t = [G|t \cdot H] \prec P_r(a,b,c).$$

Since $F_2 \prec F_t$, we only need to prove that if $F_2 \prec P_r(a,b,c)$ for some r, then $F_t \prec P_R(a,b,c)$ for some R. Suppose then F_2 is contained in the product $P_r(a,b,c)$ for some r. The idea is to find a subconfiguration of $P_r(a,b,c)$ in which there are some columns with multiplicity 1, and for the columns with multiplicity 2 or more, the multiplicity depends on r, and goes to infinity as r goes to infinity. Then we may take r to be large enough so that the multiplicity of any one column (with multiplicity of 2 or more) is larger than t.

Let x be the number of rows of F_t. Notice the following three facts, which include definitions for $E_{\mathcal{I}}$, $E_{\mathcal{I}^c}$ and $E_{\mathcal{T}}$.

$$\begin{aligned} E_{\mathcal{I}}(x,r) &:= [(r-x)\cdot \mathbf{0}_x \mid \mathcal{I}_x] \prec \mathcal{I}_r \\ E_{\mathcal{I}^c}(x,r) &:= [(r-x)\cdot \mathbf{1}_x \mid \mathcal{I}^c_x] \prec \mathcal{I}^c_r \\ E_{\mathcal{T}}(x,r) &:= \left\lfloor \frac{r}{x} \right\rfloor \cdot \mathcal{T}_x \prec \mathcal{T}_r. \end{aligned}$$

The first and second facts are easy to see; just take any subset of x rows from $\mathcal{I}_r$ or $\mathcal{I}^c_r$. The third statement is true by taking the $\lfloor r/x \rfloor$-th row of $\mathcal{T}_r$, the $2\lfloor r/x \rfloor$-th row of $\mathcal{T}_r$, etc., up to the $x\lfloor r/x \rfloor$-th row. For example, if $r = 5$ and $x = 2$, we may take the second and fourth row from $\mathcal{T}_5$:

$$\mathcal{T}_5 = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \implies \mathcal{T}_5|_{\{2,4\}} = \begin{bmatrix} 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} = E_{\mathcal{T}}(2,5)$$

Note that in the three configurations $E_{\mathcal{I}}(x,r), E_{\mathcal{I}^c}(x,r)$ and $E_{\mathcal{T}}(x,r)$, we have that there are some columns of multiplicity 1 and there are some columns for which their multiplicity can be made as large as we wish by making r large. Formally, let $E(x,r)$ be one of $E_{\mathcal{I}}(x,r)$ or $E_{\mathcal{I}^c}(x,r)$ or $E_{\mathcal{T}}(x,r)$. We have that for every x-rowed column α there are three possibilities: either $\lambda(\alpha, E(x,r)) = 0$ for all r, or $\lambda(\alpha, E(x,r)) = 1$ for all r, or $\lim_{r\to\infty} \lambda(\alpha, E(x,r)) = \infty$.

If α is a column for which $\lim_{r\to\infty} \lambda(\alpha, E(x,r)) = \infty$, we may conclude that there is an R for which $\lambda(\alpha, E(x,r)) \geq t$ for every $r \geq R$.

Since F_2 is contained in $P_r(a,b,c)$ for some r, the columns in H will have multiplicity at least 2 in some subset of the rows of $P_r(a,b,c)$. We see that F_t is also a subconfiguration of $P_R(a,b,c)$. □

3 Proof of the Main Theorem

We now prove the main theorem 2.1.

Proof. First we prove that the decision problem has a certificate which can be checked in polynomial time. A certificate that indeed $X(F) < k$ would have to be a proof that $F \prec P_r(a,b,c)$ for each triple a,b,c for which $a+b+c = k$. Note that there are at most a quadratic (with respect to the number of rows) number of a,b,c's which satisfy the equation, since the question has a trivial "yes" answer when k is more than the number of rows (Proposition 2.2).

Given F and A configurations, one can easily construct a certificate that a configuration F is indeed a subconfiguration of a configuration A: explicitly state which permutation of F appears in exactly which rows and columns of A. For the case $A = P_r(a,b,c)$, a certificate only needs to specify which rows of F go inside which factors, so at most a quadratic number of these certificates-of-being-a-subconfiguration suffice.

We now prove that finding $X(F)$ is NP-hard. Suppose there existed some polynomial-time algorithm that finds $X(F)$ given F. We shall prove that there would then exist a polynomial time algorithm for GRAPH COLORING. Suppose we are given a graph G and we wish to find the minimum number of colors for which there exists a good coloring of the graph. We may assume that no isolated vertices exist.

The idea is to construct a 3-part matrix $F(G)$ in which the first two parts ensure there is no $\mathcal{T}$ or $\mathcal{I}^c$ in a maximum product of the form $P_r(a,b,c)$ with no subconfiguration $F(G)$, and the last part is constructed so that a partition into $\mathcal{I}'s$ produces a partition of the vertices of the graph into independent sets and vice-versa.

Suppose G has n vertices and e edges. Let M be a large number with $M \geq n+2$ and let S be the incidence matrix of G (*i.e.*, the edges of G are encoded as columns with two

1's corresponding to the vertices that belong to the edge). Construct the following simple matrix:

$$F(G) := \left[\begin{array}{c|c} 1 & \mathbf{1}_M \mathcal{I}_M^c \\ \hline 1 & \mathcal{T}_M \\ \hline S & 0 \end{array}\right].$$

Clearly we can construct $F(G)$ in polynomial time (with respect to the number of vertices of G). We prove now that we have $\chi(G) = X(F) - 2M + 1$, which in turn would yield a polynomial time algorithm for GRAPH COLORING, provided we had a polynomial time algorithm for finding $X(F)$.

Now, let us study the possibilities for a product of type $P_r(a, b, c)$ that does not have $F(G)$ as a subconfiguration for any r. If $b \neq 0$, then we could place all of $[1|\mathcal{I}_M^c]$ in the $\mathcal{I}^c$ part of $P_r(a, b, c)$, so $a + b + c$ would be at most $1 + M + n$ (using Proposition 2.2). The same conclusion holds when $c \neq 0$. But if we let $b = c = 0$, $P_r(a, 0, 0)$ is just a product of $\mathcal{I}$'s, so let us calculate how many $\mathcal{I}'s$ we can multiply together and still not create a subconfiguration $F(G)$.

In order for $F(G)$ to be a part of a product of $\mathcal{I}'s$, every row of $[1|\mathcal{I}_M^c]$ and $[1|\mathcal{T}_M]$ must be in a separate factor $\mathcal{I}$, since there is no $\begin{bmatrix}1\\1\end{bmatrix}$ in $\mathcal{I}$ (and also separate from the rows of $[S|0]$, since we are assuming G has no isolated vertices).

Then two rows of the $[S|0]$ part can be in the same $\mathcal{I}$ if and only if there is no $\begin{bmatrix}1\\1\end{bmatrix}$ in those two rows, which, in terms of the graph, means there is no edge between those two vertices. On the other hand, a product of $\chi(G)$ identity matrices contains the incidence matrix of a complete $\chi(G)$-multipartite graph. In other words, partitioning $[S|0]$ into $\mathcal{I}$'s is equivalent to partitioning the vertices of G into independent sets. So if the graph G cannot be colored with $\chi(G) - 1$ colors and this is the maximum, this means that $X(F) = a = 2M + \chi(G) - 1 \geq n + M + 1$. Then $\chi(G) = X(F(G)) - 2M + 1$. □

Acknowledgements The author would like to thank the anonymous referees for their invaluable comments that doubtlessly improved the presentation of the paper.

References

[1] R. Anstee, A survey of forbidden configuration results, `http://www.combinatorics.org/ojs/index.php/eljc/article/view/DS20`.

[2] R. Anstee, R. Ferguson and A. Sali, Small forbidden configurations II, *Electronic Journal of Combinatorics* **8** (2001), R4 25pp.

[3] R. Anstee and B. Fleming, Two refinements of the bound of Sauer, Perles and Shelah and Vapnik and Chervonenkis, *Discrete Mathematics* **310** (2010), 3318–3323.

[4] R. Anstee, J. Griggs and A. Sali, Small forbidden configurations, *Graphs and Combinatorics* **13** (1997), 97–118.

[5] R. Anstee and P. Keevash, Pairwise intersections and forbidden configurations, *European Journal of Combinatorics* **27** (2006), 1235–1248.

[6] R. Anstee, M. Raggi and A. Sali, Evidence for a forbidden configuration conjecture: One more case solved, *Discrete Mathematics* **312** (2012), 2720 – 2729, doi:10.1016/j.disc.2012.02.020.

[7] R. Anstee, M. Raggi and A. Sali, Forbidden configurations: Boundary cases, *European Journal of Combinatorics* (2014), 51.

[8] R. Anstee and A. Sali, Small forbidden configurations IV, *Combinatorica* **25** (2005), 503–518.

[9] D. de Caen and Z. Füredi, The maximum size of 3-uniform hypergraphs not containing a Fano plane, *Journal of Combinatorial Theory, Series B* **78** (2000), 274–276.

[10] P. Erdős and M. Simonovits, A limit theorem in graph theory., *Studia Scientiarum Mathematicarum Hungarica* **1** (1966), 51–57.

[11] P. Erdős and A. Stone, On the structure of linear graphs, *Bulletin of the American Mathematical Society* **52** (1946), 1089–1091.

[12] Z. Füredi, Turán type problems, *Surveys in Combinatorics (Proc. of the 13th British Combinatorial Conference), ed. A.D. Keedwell, Cambridge Univ. Press. London Math. Soc. Lecture Note Series* **166** (1991), 253–300.

[13] O. Pikhurko, An exact bound for Turán result for the generalized triangle, *Combinatorica* **28** (2008), 187–208.

[14] M. Raggi, *Forbidden Configurations*, Ph.D. thesis, University of British Columbia, 2011.

Also available at http://amc-journal.eu
ISSN 1855-3966 (printed edn.), ISSN 1855-3974 (electronic edn.)
ARS MATHEMATICA CONTEMPORANEA 10 (2016) 9–18

Counting even and odd restricted permutations

Vladimir Baltić
University of Belgrade, Faculty of Organizational Sciences, Jove Ilića 154, 11000 Belgrade, Serbia

Dragan Stevanović *
University of Primorska, Institute Andrej Marušič, Muzejski trg 2, 6000 Koper, Slovenia
and
Mathematical Institute of the Serbian Academy of Science and Arts, Knez Mihajlova 36, 11000 Belgrade, Serbia

Received 5 May 2014, accepted 15 May 2014, published online 24 January 2015

Abstract

Let p be a *permutation* of the set $\mathbb{N}_n = \{1, 2, \ldots, n\}$. We introduce techniques for counting $N(n; k, r, I; \pi)$, the number of even or odd restricted permutations of $\mathbb{N}_n$ satisfying the conditions $-k \leq p(i) - i \leq r$ (for arbitrary natural numbers k and r) and $p(i) - i \notin I$ (for some set I) and $\pi = 0$ for even permutations and $\pi = 1$ for odd permutations.

Keywords: Even and odd restricted permutations, exact enumeration, recurrences, permanents.

Math. Subj. Class.: 05A15, 05A05, 11B37, 15A15

1 Introduction

Let p be a *permutation* of the set $\mathbb{N}_n = \{1, 2, \ldots, n\}$. So, $p(i)$ refers to the value taken by the function p when evaluated at a point i. A class of permutations in which the positions of the marks after the permutation are restricted can be specified by an $n \times n$ $(0, 1)$-matrix $A = (a_{ij})$ in which:

$$a_{ij} = \begin{cases} 1, & \text{if the mark } j \text{ is permitted to occupy the } i\text{-th place;} \\ 0, & \text{otherwise.} \end{cases}$$

The following result is a well known fact on the number of restricted permutations.

*The second author was supported by Research Programme P1-0285 and Research Project J1-4021 of the Slovenian Research Agency and Research Grant ON174033 of the Ministry of Education and Science of Serbia.

E-mail addresses: baltic@matf.bg.ac.rs (Vladimir Baltić), dragance106@yahoo.com (Dragan Stevanović)

Theorem 1.1 ([1]). *The number of restricted permutations is given by the permanent of a square matrix A:*

$$\operatorname{per} A = \sum_{p \in S_n} a_{1p(1)} a_{2p(2)} \cdots a_{np(n)},$$

where p runs through the set S_n of all permutations of $\mathbb{N}_n$. □

Next, we will define strongly and weakly restricted permutations (for more informations see [13]).

Definition 1.2. In *strongly restricted permutations* of $\mathbb{N}_n$, the number $r_i = \sum_{j=1}^{n} a_{ij}$ is uniformly small, i.e., $r_i \leq K$ $(i = 1, 2, \ldots, n)$, where K is an integer independent of n. In *weakly restricted permutations*, $n - r_i$ is uniformly small.

Let us briefly overview historical development of the topic of restricted permutations. Probably the most well known example is the derangement problem or "le Problème des Rencontres" (see [1] or [6]). Most of the restricted permutations considered in current literature deal with pattern avoidance. For surveys of such studies, see [3] or [9]. For a related topic of pattern avoidance in compositions and words see [5].

Detailed introduction to weakly restricted permutations can be found in [1]. A general method of enumeration of permutations with restricted positions was developed by Kaplansky and Riordan in a series of papers (they developed the theory of rook polynomials for these purposes—see [6], [7], [8], [18]). Lagrange, Lehmer, Mendelsohn, Tomescu and Stanley ([12, 13, 15, 16, 20, 19]) studied particular types of strongly restricted permutations satisfying the condition $|p(i) - i| \leq d$, where d is 1, 2, or 3 (more information on their work can be found in [1]).

Lehmer [13] classified some sets of strongly restricted permutations. The first author showed in [1] how to handle *all* five types of Lehmer's permutations. For the number of restricted permutations in a circular case the following is known: Stanley [19, Example 4.7.7] explored type $k = 2$ with the transfer-matrix method, Baltić [2] used finite state automata for type $k = 2$, and Li et al. [14] explored the $k = 3$ by expanding permanents.

An explicit technique for creating a system of the recurrence equations was given in [1], based on a simple mapping φ from combinations of $\mathbb{N}_{k+r+1}$ and some crucial differences between the transfer-matrix Method and the newly proposed technique were given.

Krafft and Schaefer in [11] find the closed formula for the strongly restricted permutations of the set $\mathbb{N}_n$ satisfying the condition $|p(i) - i| \leq k$, where $k + 2 \leq n \leq 2k + 2$. Panholzer [17] and Kløve [10] made progress in symmetric cases (Panholzer used finite state automata, while Kløve used modified transfer-matrix method based on expanding permanent) and they found the asymptotic expansion and gave bounds for the denominator of corresponding generating functions.

Here we pursue the more general, asymmetric cases and the cases where more numbers are forbidden than in the ordinary derangements for even and odd strongly restricted permutations. Our method determines the number of restricted permutations that are even, and the number of restricted permutations that are odd.

In Section 2 we introduce a general technique for counting $N(n; k, r, I; \pi)$, the number of even or odd restricted permutations ($N(n; k, r, I; \pi)$ is defined in abstract). In Section 3 we illustrate it with several examples. Using a program that implements our technique, we have contributed about a hundred sequences to the Sloane's online encyclopedia of integer sequences [21].

2 Counting $N(n; k, r, I; \pi)$

We established the connection between the number of restricted permutations and the permanent function of a matrix A, $\operatorname{per} A$, in Theorem 1 from the introduction. The Laplace expansion of the permanent function (this is the same as for the determinant function) is computationally inefficient for high dimension because for $n \times n$ matrices, the computational effort scales with $n!$. Therefore, the Laplace expansion is not suitable for large n. However, the matrices obtained in the Laplace expansions for restricted permutations have the regular structure, so called band matrices (a band matrix is a sparse matrix whose non-zero entries are confined to a diagonal band, comprising the main diagonal and zero or more diagonals on either side), and their expansions can be reduced to a system of linear recurrence equations.

We present a general technique for counting $N(n; k, r, I; \pi)$, the number of even or odd ($\pi = 0$ for even permutations and $\pi = 1$ for odd permutations) restricted permutations satisfying the conditions $-k \leq p(i) - i \leq r$ and $p(i) - i \notin I$ for all $i \in \mathbb{N}_n$, where $k \leq r < n$, and I is a fixed subset of the set $\{-k+1, -k+2, \ldots, r-1\}$. Assume that I contains x elements, $|I| = x$. Our technique proceeds in six steps:

1. Create $\mathcal{C}$, a set of all $(k+1)$-element combinations of the set $\mathbb{N}_{k+r+1}$ containing element $k+r+1$.
2. Create $\mathcal{D}$, a set of all ordered pairs $D = (C, \pi)$, where $C \in \mathcal{C}$ and $\pi \in \{0, 1\}$.
3. Introduce an integer sequence $a_D(n)$ for each ordered pair $D \in \mathcal{D}$.
4. Apply the mapping φ (defined below) to each ordered pair.
5. Create a system of linear recurrence equations (later we will see that these equations correspond to the Laplace expansion of a permanent of the matrix A):

$$a_D(n) = \sum_{D' \in \varphi(D)} a_{D'}(n-1).$$

6. Solve the system to obtain equations $N(n; k, r, I; 0) = a_{((r+1, r+2, \ldots, r+k+1), 0)}(n)$ and $N(n; k, r, I; 1) = a_{((r+1, r+2, \ldots, r+k+1), 1)}(n)$.

We next describe these steps in detail and then prove that $N(n; k, r, I; \pi)$ is indeed equal to $a_{((r+1, r+2, \ldots, r+k+1), \pi)}(n)$.

Definition 2.1. Let $\mathcal{C}$ denote a set of all combinations with $k+1$ elements of the set $\mathbb{N}_{k+r+1}$, which contain $k+r+1$. We represent these combinations as strictly increasing ordered $(k+1)$-tuples.

For example, all such combinations with 3 elements of the set $\mathbb{N}_5 = \{1, 2, 3, 4, 5\}$ are represented (in reverse lexicographic order) by:

$$(3,4,5), \quad (2,4,5), \quad (2,3,5), \quad (1,4,5), \quad (1,3,5), \quad (1,2,5).$$

In examples we will use easier notation:

$$345, \qquad 245, \qquad 235, \qquad 145, \qquad 135, \qquad 125.$$

Definition 2.2. Let $\alpha \pm I$ denote the set $\alpha \pm I = \{\alpha \pm i \mid i \in I\}$.

We split the set $\mathcal{C}$ in two disjoint sets

$$\mathcal{C}_1 = \{C \in \mathcal{C} \mid 1 \in C\} \qquad \text{and} \qquad \mathcal{C}_2 = \{C \in \mathcal{C} \mid 1 \notin C\},$$

but we will also separate the set $\mathcal{C}_2$ into $x+1$ disjoint sets

$$\mathcal{C}_2^m = \{C \in \mathcal{C}_2 \mid m \text{ elements of } C \text{ are in } r+1-I\}, \quad (m = 0, 1, \ldots, x).$$

Let $\mathcal{C}^{k+1-m}$ denote a Cartesian product $\mathcal{C}^{k+1-m} = \mathcal{C} \times \mathcal{C} \times \cdots \times \mathcal{C}$, where $\mathcal{C}$ appears $(k+1-m)$ times. Let B denote the set $B = \{0, 1\}$.

We define the set of ordered pairs $\mathcal{D} = \{(C, \pi) \mid C \in \mathcal{C}, \pi \in B\}$ and same as in the case of $\mathcal{C}$ we will divide it into disjoint sets:

$$\mathcal{D}_1 = \{(C, \pi) \mid C \in \mathcal{C}_1, \pi \in B\}, \qquad \mathcal{D}_2 = \{(C, \pi) \mid C \in \mathcal{C}_2, \pi \in B\},$$

$$\mathcal{D}_2^m = \{(C, \pi) \mid C \in \mathcal{C}_2^m, \pi \in B\}, \quad (m = 0, 1, \ldots, x).$$

For each $D \in \mathcal{D}_2$ we define ordered $(k+1)$-tuple

$$SD = (D_1, D_2, \ldots, D_k, D_{k+1})$$

in the following manner. We get each of the combinations $C_i \in \mathcal{C}$ from the initial combination $C = (c_1, c_2, \ldots, c_k, c_{k+1})$ (pay attention that $D \in \mathcal{D}$ is $D = (C, \pi)$) by deleting c_i, decreasing all other coordinates by 1, shifting all coordinates with bigger index to one place left and putting $k+r+1$ at the end:

$$C_i = (c_1 - 1, \ldots, c_{i-1} - 1, c_{i+1} - 1, \ldots, c_{k+1} - 1, k+r+1).$$

For the parity coordinate, we have an easier condition:

$$\pi_i = \begin{cases} \pi, & i \text{ is odd,} \\ 1 - \pi, & i \text{ is even,} \end{cases}$$

i.e. if i is odd the parity coordinate stays the same and if i is even the parity coordinate changes.

In the same way as before, we also introduce D_1 for each $D \in \mathcal{D}_1$:

$$D_1 = \big((c_2 - 1, c_3 - 1, \ldots, c_k - 1, c_{k+1} - 1, k+r+1), \pi\big)$$

(in this case the parity coordinate π stays the same).

Now, we get ordered $(k+1-m)$-tuple $SD' = \big(D'_1, D'_2, \ldots, D'_{k+1-m}\big)$ from ordered $(k+1)$-tuple $SD = (D_1, D_2, \ldots, D_k, D_{k+1})$ when we delete all ordered pairs $D_y = (C_y, \pi_y)$ corresponding to elements c_y which satisfy the condition $c_y \in r+1-I$.

Finally, we introduce the mapping

$$\varphi(D) = \begin{cases} \varphi_1(D), & D \in \mathcal{D}_1 \\ \varphi_2^m(D), & D \in \mathcal{D}_2^m, \end{cases}$$

defined by $\varphi_1 : \mathcal{D}_1 \to \mathcal{D}$ and $\varphi_2^m : \mathcal{D}_2^m \to \mathcal{D}^{k+1-m}$, for $m = 0, 1, \ldots, x$, defined by

$$\varphi_1(D) = D_1, \qquad \varphi_2^m(D) = SD'.$$

We use these mappings to find a system of $2 \cdot \binom{k+r}{k}$ linear recurrence equations (one equation per ordered pair, i.e. two equations per combination – one corresponding to even permutations and another corresponding to odd permutations): if we have $\varphi_1(D) = D'$ then we have the linear recurrence equation:

$$a_D(n+1) = a_{D'}(n)$$

and if we have $\varphi_2^m(D) = (D'_1, D'_2, \ldots, D'_{k+1-m})$ then we have the linear recurrence equation:

$$a_D(n+1) = a_{D'_1}(n) + a_{D'_2}(n) + \cdots + a_{D'_{k+1-m}}(n).$$

The initial conditions are: $a_{((r+1,r+2,\ldots,r+k+1),0)}(0) = 1$ and $a_D(0) = 0$ for all $D \neq ((r+1, r+2, \ldots, r+k+1), 0)$.

This system can be easily solved, for example by using the standard method based on generating functions. From this system of linear recurrence equations we are able to get a linear recurrence equation and a generating function for $N(n; k, r, I; \pi)$. We will prove that $N(n; k, r, I; 0) = a_{((r+1,r+2,\ldots,r+k+1),0)}(n)$ and $N(n; k, r, I; 1) = a_{((r+1,r+2,\ldots,r+k+1),1)}(n)$. Thus, from the matrix of this system, S, we can find $N(n; k, r, I; \pi)$ as the element in the first row and the first column of the matrix S^n, i.e., the number of the closed paths in the digraph G whose adjacency matrix is S (this observation is important because we can apply the Transfer matrix method to the matrix S). We apply this observation to determine the computational complexity of our technique: S^n can be computed with repeated squaring [4] in $O(\log_2 n)$ operations. Hence, our technique evaluates the number of restricted permutations more efficiently than the straightforward techniques of filtering permutations or expanding the permanent $\operatorname{per} A$.

All the generating functions that we derive using our technique are rational. We have a system of $2 \cdot \binom{k+r}{k}$ linear recurrence equations which leads us to the upper bound for the degree d of the denominator polynomial: $d \leq 2 \cdot \binom{k+r}{k}$. It is sufficient to compute a finite number of values, in particular $2 \cdot \binom{k+r}{k}$ of them, to find the generating function.

Theorem 2.3. *For even permutations* $N(n; k, r, I; 0) = a_{((r+1,r+2,\ldots,r+k+1),0)}(n)$ *and for odd permutations* $N(n; k, r, I; 1) = a_{((r+1,r+2,\ldots,r+k+1),1)}(n)$.

Proof. We establish the correspondence between combination $C = (c_1, c_2, \ldots, c_k) \in \mathcal{C}$ and the specific matrix $M_C = f(C)$. We introduce a set $\mathcal{M}_t$ (for a fixed t) of matrices M_C that correspond to the sequences $a_{D_0}(n)$ and $a_{D_1}(n)$, where $D_0 = (C, 0)$ and $D_1 = (C, 1)$.

Let matrix $M_C = (m_{ij})$ satisfies the following conditions:

1) the first $k+1$ rows is definied by:

   for $i = 1, 2, \ldots, k+1$, $m_{ij} = \begin{cases} 1, & j + r - c_i \notin I \\ 0, & j + r - c_i \in I \end{cases}$ for $j = 1, 2, \ldots, c_i$ and $m_{ij} = 0$ for $j > c_i$;

2) elements in the last $t - (k+1)$ rows satisfy: $m_{ij} = 1$ for $-k \leq j - i \leq r$, $j - i \notin I$ and $m_{ij} = 0$ otherwise.

Denote by $\mathcal{M}_t$ the set of all $t \times t$ ($t > r$) matrices M_C for $C \in \mathcal{C}$.

From the matrix $M_C \in \mathcal{M}_t$, we can determine the corresponding combination $C = (c_1, c_2, \ldots, c_k) \in \mathcal{C}$: let c_i denotes the column of the last one in the i-th row of the matrix M, $i = 1, 2, \ldots, k+1$.

So, the function $f : \mathcal{C} \to \mathcal{M}_t$, defined by $f(C) = M_C$ is a bijection.

We associate an $n \times n$ matrix $A = (a_{ij})$ defined by:

$$a_{ij} = \begin{cases} 1, & \text{if} \quad -k \le j - i \le r, \ j - i \notin I, \\ 0, & \text{otherwise} \end{cases}$$

with the strongly restricted permutations satisfying $-k \le p(i) - i \le r$ and $p(i) - i \notin I$. As stated in the introduction, the number of all permutations (even and odd) satisfying $-k \le p(i) - i \le r$ and $p(i) - i \notin I$ is equal to $\operatorname{per} A$. Notice that $A \in \mathcal{M}_n$ with $c_i = r + i$, where $1 \le i \le k + 1$, and thus the combination corresponding to A is $(r + 1, r + 2, \dots, r + k + 1)$.

We next observe that the recurrence equations from step 5. (see page 11) correspond to the expansion of the permanent of matrices from $\mathcal{M}_t$ by the first row (φ_1) or by the first column (in cases of all of φ_2^m; note that when we skip an element c_y, it corresponds to a zero element in the first column). During this expansion we need to take care about the parity of the permutation under construction.

First, note that at each step of construction determines the position of the smallest of the remaining elements of the permutation. Let q denote the number of already used elements in the construction of the restricted permutation. Define a monotonically increasing sequence w of positions in the permutation which have not yet been assigned values: $w = (w_1, w_2, \dots, w_{n-q})$, where $w_1 < w_2 < \cdots < w_{n-q}$.

If we make an expansion by the first row (we have one in the first column), it corresponds to $p(w_1) = q + 1$ and the parity of the permutation under construction doesn't change because we haven't got any new inversions.

If we make an expansion by the first column and if we have 1 in the i-th row (i.e. at the position $(i, 1)$ in the matrix A is 1), it corresponds to $p(w_i) = q + 1$. There are $i - 1$ numbers: $p(w_1), p(w_2), \dots, p(w_{i-1})$ which made new inversions with $p(w_i) = q + 1$, because all of them have not been assigned yet, so they are all greater than $q + 1$. So, the parity of the permutation under construction depends on the parity of i:

- if i is even then there are odd number ($i - 1$) inversions, so we need to change the parity of the permutation under construction, $\pi' = 1 - \pi$;
- if i is odd then there are even number ($i - 1$) inversions, so we don't need to change the parity of the permutation under construction, $\pi' = \pi$.

These observations lead to the main conclusions:

$$N(n; k, r, I; 0) = a_{((r+1,\dots,r+k+1),0)}(n)$$
$$N(n; k, r, I; 1) = a_{((r+1,r+2,\dots,r+k+1),1)}(n).$$

□

3 Examples

We illustrate the technique from the previous section on two examples.

Example 3.1. We find the number of even (odd) permutations of the set $\mathbb{N}_n$, satisfying the condition $-1 \le p(i) - i \le 1$ for all $i \in \mathbb{N}_n$. It is usually referred to as a permutation of length n within distance 1.

In this case we have $k = r = 1$, i.e. $k + r + 1 = 3$ and $\mathcal{C} = \{23, 13\}$.
$\varphi_2(23, 0) = \{(23, 0), (13, 1)\}$, $\varphi_1(13, 0) = \{(23, 0)\}$, $\varphi_2(23, 1) = \{(23, 1), (13, 0)\}$, $\varphi_1(13, 1) = \{(23, 1)\}$, from which we get the system of linear recurrence equations:

$$\begin{array}{rcl} a_{(23,0)}(n+1) & = & a_{(23,0)}(n) + a_{(13,1)}(n), \\ a_{(13,0)}(n+1) & = & a_{(23,0)}(n), \\ a_{(23,1)}(n+1) & = & a_{(23,1)}(n) + a_{(13,0)}(n), \\ a_{(13,1)}(n+1) & = & a_{(23,1)}(n), \end{array}$$

with the initial conditions $a_{(23,0)}(0) = 1$, $a_{(13,0)}(0) = 0$, $a_{(23,1)}(0) = 0$, $a_{(13,1)}(0) = 0$. If we substitute $a_{(23,0)}(n) = a_n$, $a_{(13,0)}(n) = b_n$, $a_{(23,1)}(n) = c_n$ and $a_{(13,1)}(n) = d_n$ we have a simpler form:

$$a_{n+1} = a_n + d_n, \qquad b_{n+1} = a_n, \qquad c_{n+1} = c_n + b_n, \qquad d_{n+1} = c_n.$$

The initial conditions are $a_0 = 1$, $b_0 = c_0 = d_0 = 0$.

For a sequence which is denoted by a lower case letter we will denote the corresponding generating function by the same upper case letter ($a_n \leftrightarrow A(z)$, $b_n \leftrightarrow B(z)$, and so on). We find the following system of linear equations $\big($variables are $A(z), B(z), C(z), D(z)\big)$:

$$\frac{A(z) - 1}{z} = A(z) + D(z), \quad \frac{B(z)}{z} = A(z), \quad \frac{C(z)}{z} = C(z) + B(z), \quad \frac{D(z)}{z} = C(z)$$

and part of its solution that we are interested in is:

$$A(z) = \frac{1 - z}{1 - 2z + z^2 - z^4}, \qquad C(z) = \frac{z^2}{1 - 2z + z^2 - z^4}.$$

From the denominator of these generating functions $1 - 2z + z^2 - z^4$, we can find the linear recurrence equations $a_n = 2a_{n-1} - a_{n-2} + a_{n-4}$ and $c_n = 2c_{n-1} - c_{n-2} + c_{n-4}$.

When we solve these equations we find the general terms of these sequences:

$$a_n = \frac{1}{2}\left(F_{n+1} + x_n\right), \qquad c_n = \frac{1}{2}\left(F_{n+1} - x_n\right),$$

where F_n denotes n-th Fibonacci number ($F_1 = F_2 = 1$, $F_{n+1} = F_n + F_{n-1}$; A000045 at [21]), and $x_n = \cos\frac{n\pi}{3} + \frac{1}{\sqrt{3}}\sin\frac{n\pi}{3}$ (A010892 at [21]).

The number of even permutations, a_n, and odd permutations, c_n, both satisfying the condition $|p(i) - i| \le 1$, for all $i \in \mathbb{N}_n$ is determined by previous formulae or by their generating functions $A(z)$ and $C(z)$:

n	0	1	2	3	4	5	6	7	8	9	10	...
a_n	1	1	1	1	2	4	7	11	17	27	44	...
c_n	0	0	1	2	3	4	6	10	17	28	45	...

These sequences are A005252 and A024490 at [21]. ■

Example 3.2. We find the number of even (odd) permutations of the set $\mathbb{N}_n$, satisfying the condition $p(i) - i \in \{-2, 0, 2\}$.

In this case we have $k = r = 2$, i.e. $k + r + 1 = 5$, set $I = \{-1, 1\}$, which implies $(r + 1 - I) = \{2, 4\}$ and $\mathcal{C} = \{345, 245, 235, 145, 135, 125\}$, which is separated into sets:

$$\mathcal{C}_1 = \{145, 135, 125\}, \qquad \mathcal{C}_2^0 = \emptyset, \qquad \mathcal{C}_2^1 = \{345, 235\}, \qquad \mathcal{C}_2^2 = \{245\}.$$

$$\begin{array}{ll}
\varphi_2^1(345,0) = \{(345,0),(235,0)\}, & \varphi_2^1(345,1) = \{(345,1),(235,1)\},\\
\varphi_2^2(245,0) = \{(135,0)\}, & \varphi_2^2(245,1) = \{(135,1)\},\\
\varphi_2^1(235,0) = \{(145,1),(125,0)\}, & \varphi_2^1(235,1) = \{(145,0),(125,1)\}\\
\varphi_1(145,0) = \{(345,0)\}, & \varphi_1(145,1) = \{(345,1)\},\\
\varphi_1(135,0) = \{(245,0)\}, & \varphi_1(135,1) = \{(245,1)\},\\
\varphi_1(125,0) = \{(145,0)\}, & \varphi_1(125,1) = \{(145,1)\}.
\end{array}$$

If we substitute $a_{(345,0)}(n) = a_n$, $a_{(245,0)}(n) = b_n$, $a_{(235,0)}(n) = c_n$, $a_{(145,0)}(n) = d_n$, $a_{(135,0)}(n) = e_n$, $a_{(125,0)}(n) = f_n$, $a_{(345,1)}(n) = g_n$, $a_{(245,1)}(n) = h_n$, $a_{(235,1)}(n) = i_n$, $a_{(145,1)}(n) = j_n$, $a_{(135,1)}(n) = k_n$ and $a_{(125,1)}(n) = \ell_n$ we get the system of linear recurrence equations:

$$\begin{array}{ll}
a_{n+1} = a_n + c_n, & g_{n+1} = g_n + i_n,\\
b_{n+1} = e_n, & h_{n+1} = k_n,\\
c_{n+1} = j_n + f_n, & i_{n+1} = d_n + \ell_n\\
d_{n+1} = a_n, & j_{n+1} = g_n,\\
e_{n+1} = b_n, & k_{n+1} = h_n,\\
f_{n+1} = d_n, & \ell_{n+1} = j_n,
\end{array}$$

with the initial conditions $a_0 = 1$ and $b_0 = c_0 = \cdots = \ell_0 = 0$.

From this system we find the generating functions:

$A(z) = \frac{1-z-z^4}{1-2z+z^2-2z^4+2z^5-z^6+z^8}$ and $G(z) = \frac{z^3}{1-2z+z^2-2z^4+2z^5-z^6+z^8}$.

From the denominator of these generating functions

$$1 - 2z + z^2 - 2z^4 + 2z^5 - z^6 + z^8 = (1 - z)(1 + z)(1 + z^2)(1 - z + z^2)(1 - z - z^2),$$

we find the linear recurrence equation $a_n = 2a_{n-1} - a_{n-2} + 2a_{n-4} - 2a_{n-5} + a_{n-6} - a_{n-8}$ and same for g_n.

When we solve this equation we find the general terms of these sequences:

$$a_n = \frac{1}{10}\left(L_{n+2} + y_n + z_n\right), \qquad g_n = \frac{1}{10}\left(L_{n+2} + y_n - z_n\right),$$

where L_n denotes n-th Lucas number ($L_1 = 1$, $L_2 = 3$, $L_{n+1} = L_n + L_{n-1}$; A000032 and A000204 at [21]), $y_n = 2\cos\frac{n\pi}{2} + \sin\frac{n\pi}{2}$ and $z_n = 5$, if n is congruent to 0, 1 or 2 modulo 6, and $z_n = 0$, if n is congruent to 3, 4 or 5 modulo 6.

The number of even permutations, a_n, and odd permutations, g_n, both satisfying the conditions $|p(i) - i| \leq 2$ and $p(i) - i \neq -1, 1$ is determined by previous formulae or by their generating functions $A(z)$ and $G(z)$:

n	0	1	2	3	4	5	6	7	8	9	10	...
a_n	1	1	1	1	2	3	5	8	13	20	32	...
g_n	0	0	0	1	2	3	4	7	12	20	32	...

■

4 Concluding remarks

We have developed a technique for generating a system of linear recurrence equations that enumerate the even and the odd strongly restricted permutations. In some cases, using the digraph corresponding to the matrix of the system we can establish a connection between restricted permutations and restricted compositions. Using a program that implements this technique, we have contributed 96 sequences, A241975–A242070, to the Online encyclopedia of integer sequences [21].

We thank anonymous referees for carefully reading the manuscript and helpful suggestions which led to considerable improvements in presentation of results.

References

[1] V. Baltić, On the number of certain types of strongly restricted permutations, *Appl. Anal. Discr. Math.* **4** (2010), No **1**, 119-135.

[2] V. Baltić, Applications of the finite state automata for counting restricted permutations and variations, *Yugosl. J. Oper. Res.* **22** (2012), No **2**, 183-198.

[3] M. Bona, *Combinatorics of permutations*, Chapman & Hall/CRC, 2004.

[4] T. H. Cormen, C. E. Leiserson, R. L. Rivest and C. Stein, *Introduction to algorithms*, The MIT Press, 2001.

[5] S. Heubach and T. Mansour, *Combinatorics of compositions and words*, Chapman & Hall/CRC an imprint of Taylor & Francis LLC, 2009.

[6] I. Kaplansky, On a generalization of the “problème des rencontres”, *Amer. Math. Monthly.* **46** (1939), 159–161.

[7] I. Kaplansky, Solution of the “problème des ménages”, *Bull. Amer. Math. Soc.* **49** (1943), 784–785.

[8] I. Kaplansky and J. Riordan, The problem of the rooks and its applications, *Duke Math. J.* **13** (1946), 259–268.

[9] S. Kitaev, *Patterns in permutations and words*, Springer Verlag, 2011.

[10] T. Kløve, Generating functions for the number of permutations with limited displacement, *Electron. J. Comb.* **16** (2009), #R104.

[11] O. Krafft and M. Schaefer, On the number of permutations within a given distance, *Fibonacci Quart.* **40** (2002), 429–434.

[12] M. R. Lagrange, Quelques résultats dans la métrique des permutations, *Ann. Sci. Ecole Norm. S.* **79** (1962), 199–241.

[13] D. H. Lehmer, Permutations with strongly restricted displacements, *Combinatorial Theory and its appl II (Proc. Colloq., Balatonfured, 1969)*, North-Holland, Amsterdam, 1970, 755-770.

[14] L. Li, T. Tomoda, S. Midorikawa and T. Horibata, A Counting Formula for the Number of Permutations with Constraints, *Z. angew. Math. Mech.* **76** (1996), S3, 495–496.

[15] N. S. Mendelsohn, The asymptotic series for a certain class of permutation problems, *Can. J. Math.* **8** (1956), 234–244.

[16] N. S. Mendelsohn, Permutations with confined displacements, *Can. Math. Bull.* V.**4** (1961), 29–38.

[17] A. Panholzer, On generalized Fibonacci Permutations, *J. Inf. Opt. Sci.* **24** (2003), 591–610.

[18] J. Riordan, *An Introduction to Combinatorial Analysis*, John Wiley & Sons, Inc, 1958.

[19] R. P. Stanley, *Enumerative Combinatorics, Vol. 1*, Wadsworth, 1999.

[20] I. Tomescu, *Problems in combinatorics and graph theory*, Wiley–Interscience Series in Discrete Mathematics, New York, 1985, 60;292–295.

[21] The On-Line Encyclopedia of Integer Sequences, published electronically at `http://oeis.org/`

Also available at http://amc-journal.eu
ISSN 1855-3966 (printed edn.), ISSN 1855-3974 (electronic edn.)
ARS MATHEMATICA CONTEMPORANEA 10 (2016) 19–29

Squashing maximum packings of 6-cycles into maximum packings of triples

Curt C. Lindner
Department of Mathematics and Statistics, Auburn University, Auburn, AL 36849-5307, U.S.A.

Giovanni Lo Faro *
Dipartimento di Matematica e Informatica, Università di Messina, 98166 Messina, Italia

Antoinette Tripodi †
Dipartimento di Matematica e Informatica, Università di Messina, 98166 Messina, Italia

Received 21 July 2014, accepted 28 September 2014, published online 24 January 2015

Abstract

A 6-cycle is said to be squashed if we identify a pair of opposite vertices and name one of them with the other (and thereby turning the 6-cycle into a pair of triples with a common vertex). The squashing problem for 6-cycle systems was introduced by C. C. Lindner, M. Meszka and A. Rosa and completely solved by determining the spectrum. In this paper, by employing PBD and GDD-constructions and filling techniques, we extend this result by squashing maximum packings of K_n with 6-cycles into maximum packings of K_n with triples. More specifically, we establish that for each $n \geq 6$, there is a max packing of K_n with 6-cycles that can be squashed into a maximum packing of K_n with triples.

Keywords: Maximum packing with triples, maximum packing with 6-cycles.

Math. Subj. Class.: 05B07

* Supported by P.R.I.N., P.R.A. and I.N.D.A.M.(G.N.S.A.G.A.)
† Supported by P.R.I.N., P.R.A. and I.N.D.A.M.(G.N.S.A.G.A.)
E-mail addresses: lindncc@auburn.edu (Curt C. Lindner), lofaro@unime.it (Giovanni Lo Faro), atripodi@unime.it (Antoinette Tripodi)

1 Introduction

Let G be a graph. A *G-design of order n* is a pair (S, B) where B is a collection of subgraphs (*blocks*), each isomorphic to G, which partitions the edge set of the complete undirected graph K_n with vertex set S. After determining the spectrum for G-designs for different graphs G, many problems have been studied also recently (for example, see [1]-[7]).

A Steiner triple system (more simply, triple system) of order n is a G-design of order n where G is the graph K_3. It is well known that the spectrum for triple systems is precisely the set of all $n \equiv 1$ or 3 (mod 6) [9], and that if (S, T) is a triple system of order n, then $|T| = n(n-1)/6$. Similarly, a 6-cycle system of order n is a G-design of order n where G is 6-cycle. The spectrum for 6-cycle systems is precisely the set of all $n \equiv 1$ or 9 (mod 12) [15], and if (X, C) is a 6-cycle system of order n, then $|C| = n(n-1)/12$. It is worth noting that if (S, T) and (X, C) have order n, then $|T| = 2|C|$.

Given the fact that triple systems and 6-cycle systems coexist for all $n \equiv 1$ or 9 (mod 12), an obvious question to ask is: are there any connections between the two when $n \equiv 1$ or 9 (mod 12)? The answer, of course, is yes. One much studied connection is that of 2-perfect 6-cycle systems. A 6-cycle system is 2-perfect provided the collection of triples obtained by replacing each 6-cycle (a, b, c, d, e, f) with the two triples (a, c, e) and (b, d, f) is a Steiner triple system. Such systems exist for all $n \equiv 1$ or 9 (mod 12) ≥ 13 [15].

Quite recently a new connection between triple systems and 6-cycle systems has been introduced: the *squashing* of a 6-cycle system into a Steiner triple system. A definition is in order. Let (a, b, c, d, e, f) be a 6-cycle and form the following six bowties (a pair of triples with a common vertex).
If B is any one of the six bowties in Figure 1, we say that we have *squashed* (a, b, c, d, e, f) into B. So there are six different ways to squash a 6-cycle into a bowtie. If (X, C) is a 6-cycle system, $2|C| = 2n(n-1)/12 = n(n-1)/6$ is the number of triples in a Steiner triple system. Therefore it makes sense to ask the following question: what is the spectrum for 6-cycle systems that can be squashed into Steiner triple systems? In [11], a complete solution is given to this problem by constructing for every $n \equiv 1$ or 9 (mod 12) a 6-cycle system that can be squashed into a Steiner triple system.

Example 1.1. (A 6-cycle system of order 9 squashed into a triple system [11].)

(0,1,2,3,4,5)		(0,1,2)(0,4,5)
(3,6,0,2,4,1)		(3,6,0)(3,4,1)
(2,8,4,0,3,7)	SQUASH	(2,8,4)(2,3,7)
(7,0,8,6,5,1)	$\longrightarrow$	(7,0,8)(7,5,1)
(6,1,8,5,7,4)		(6,1,8)(6,7,4)
(5,2,6,7,8,3)		(5,2,6)(5,8,3)

Now if $n \equiv 3$ or 7 (mod 12) there does not exist a 6-cycle system of order n. However, there does exist a maximum packing (max packing) of K_n with 6-cycles with leave a triple (i.e., a pair (X, C) and a set L, the *leave*, where C is a collection of edge disjoint 6-cycles with verteces in X, L is the set of the edges of K_n not belonging to any 6-cycle of C and $|L|$ is as small as possible) and so the following question makes sense. Does there exist for each $n \equiv 3$ or 7 (mod 12) a max packing of K_n with 6-cycles which can be squashed into bowties so that the bowties plus the leave (a triple) form a Steiner triple system?

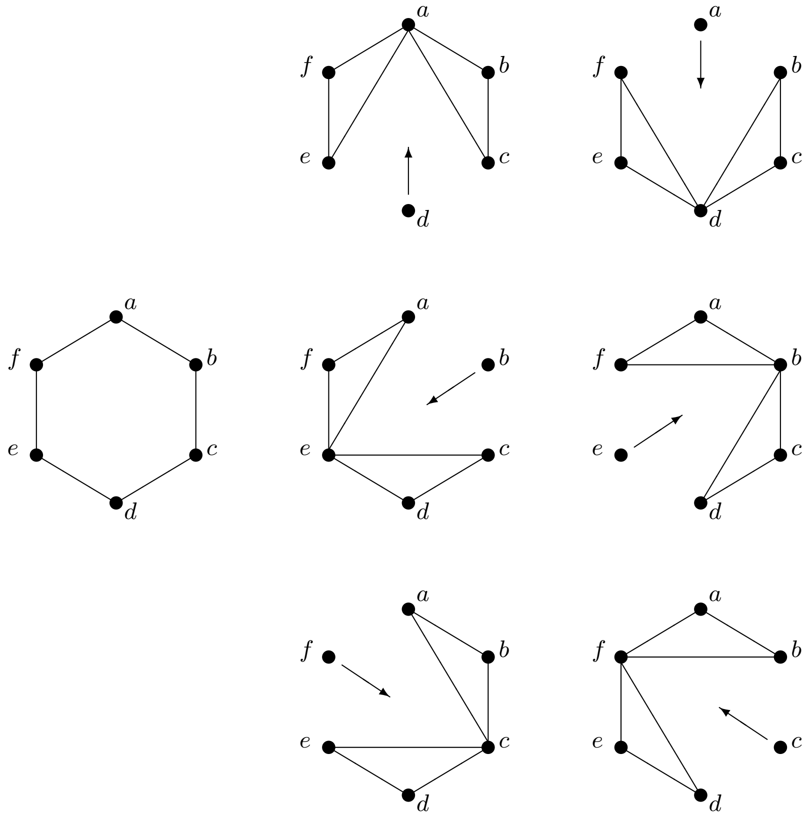


Figure 1

Example 1.2. (A max packing of K_7 squashed into a triple system [11].)

(2,3,4,5,0,1)	SQUASH	(2,1,0)(2,3,4)
(4,6,0,2,5,1)	$\longrightarrow$	(4,1,5)(4,0,6)
(5,6,2,4,0,3)		(5,3,0)(5,6,2)
(1,3,6) leave	$\longrightarrow$	(1,3,6)

The following theorem is proved in [11].

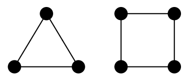
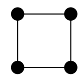
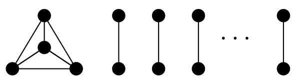
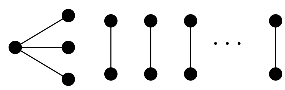
Theorem 1.3. [11] *There exists a* 6*-cycle system of every order* $n \equiv 1$ *or* 9 (mod 12) *that can be squashed into a triple system and a* 6*-cycle maximum packing that can be squashed into a triple system for every* $n \equiv 3$ *or* 7 (mod 12)*,* $n \geq 7$. □

The object of this paper is to finish off the problem of squashing maximum packings of K_n with 6-cycles into maximum packings of K_n with triples. We need to be a bit more precise.

Let (X, C) be a maximum packing of K_n with 6-cycles with leave L. In what follows, to keep the vernacular from getting out of hand, to say that C has been squashed means that the resulting collection $S(C)$ of bowties is a partial triple system.

Further, if t is a triple belonging to L and $S(C) \cup \{t\}$ is a maximum packing of K_n with triples (or a triple system), we will say that we have squashed (X, C) into a maximum packing of K_n with triples. So, for example, Example 1.2 is the squashing of a maximum packing of K_7 with 6-cycles into a triple system of order 7.

The following easy to read table gives the leaves for max packings for both 6-cycles and triples not covered by Theorem 1.3. (See [8] and [13].)

K_n	6-cycles leave	triples leave
$n \equiv 0, 2, 6, 8$ (mod 12)	1-factor	1-factor
$n \equiv 5$ (mod 12)	4-cycle	4-cycle
$n \equiv 11$ (mod 12)	 4 leaves are possible	
$n \equiv 4$ or 10 (mod 12)	 22 leaves are possible for $n \geq 16$	 tripole [13]

We remark that if $n \equiv$ 0, 2, 6, 8 or 5 (mod 12) and a 6-cycle maximum packing can be squashed, there are no triples to be added; i.e., the resulting collection of bowties is a maximum packing of K_n with triples. If $n \equiv 4$, 10 or 11 (mod 12) and a 6-cycle maximum packing can be squashed, then a triple is taken from the 6-cycle leave in order to obtain a maximum packing of K_n with triples.

2 Preliminaries

From now on to say that the 6-cycle (a, b, c, d, e, f) is squashed we will always mean that it has been squashed into the bowtie $(a, b, c)(a, e, f)$; see Figure 2.

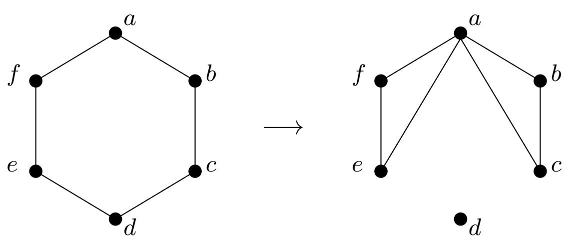


Figure 2

So, for example, in Example 1.1 we can simply list the 6-cycles (without listing the bowties they have been squashed into) and say they can be squashed into a triple system.

The following three examples are used repeatedly in what follows.

Example 2.1. (A max packing of K_6 with 6-cycles squashed into a max packing of K_6 with triples.)

$C = \{(5, 0, 1, 2, 4, 3), (2, 3, 1, 5, 4, 0)\}$, leave $L = \{(0, 3), (1, 4), (2, 5)\}$. (There are no triples in the leave.)

Example 2.2. (A max packing of K_8 with 6-cycles squashed into a max packing of K_8 with triples.)

$X = Z_4 \times Z_2$; $C = \{(0_0, 3_0, 1_1, 2_0, 3_1, 0_1), (1_0, 0_0, 2_1, 3_0, 0_1, 1_1), (2_0, 1_0, 3_1, 0_0, 1_1, 2_1), (3_0,\ 2_0, 0_1, 1_0, 2_1, 3_1)\}$, leave $L = \{(0_0, 2_0), (1_0, 3_0), (0_1, 2_1), (1_1, 3_1)\}$. (There are no triples in the leave.)

Example 2.3. (Decomposition of $K_{4,4,4}$ into 6-cycles squashed into triples.) (An obvious definition.)

$X = Z_4 \times \{1, 2, 3\}$; $C = \{(1_2, 1_3, 0_1, 0_2, 0_3, 1_1), (0_2, 2_3, 0_1, 1_2, 0_3, 2_1), (1_1, 0_2, 3_3, 0_1, 2_2, 1_3), (0_1, 3_2, 3_3, 1_1, 2_2, 0_3), (3_2, 1_1, 2_3, 1_2, 3_1, 0_3), (1_2, 2_1, 2_3, 3_2, 3_1, 3_3), (3_1, 0_2, 1_3, 2_1, 2_2, 2_3), (2_1, 3_2, 1_3, 3_1, 2_2, 3_3)\}$. (There is no leave.)

3 Basic Lemmas

With the examples of Section 2 in hand we can go to the general constructions, where we shall make use of GDDs. Let H be a set of integers and X be a set of size n; a GDD(n, H, k) is a triple (X, G, B) where G is a partition of X into subsets called *groups* of size in H, B is a set of subsets of X (called *blocks*) of size k, such that a group and a block contain at most one common point and every pair of points from distinct groups occurs in exactly one block. A PBD is a GDD$(n, \{1\}, k)$.

We break the constructions into the eight cases: 2, 6, 8; 0; 11; 4, 10; 5 (mod 12).

3.1 $n \equiv 2, 6$ and 8 (mod 12)

These are the easiest cases, so a good place to start.

$n \equiv 2$ (mod 12) Write $12k + 2 = 2(6k + 1)$ and let (X, T) be a Steiner triple system of order $6k + 1$. Let $S = X \times \{1, 2\}$ and define a collection C of 6-cycles as follows: For each triple $t = \{a, b, c\} \in T$ define a copy of Example 2.1 on $\{a, b, c\} \times \{1, 2\}$ with leave $L_t = \{(a_1, a_2), (b_1, b_2), (c_1, c_2)\}$ and put these 6-cycles in C. Then C is a max packing of K_{12k+2} with 6-cycles with leave $L = \{L_t |\ t \in T\}$. Trivially, C can be squashed into a max packing of K_{12k+2} with triples with leave L.

$n \equiv 6$ (mod 12) The case for $n = 6$ is handled with Example 2.1. So now write $12k+6 = 2(6k + 3)$ and proceed exactly as in the case $n \equiv 2$ (mod 12).

$n \equiv 8$ (mod 12) Write $12k + 8 = 2(6k + 4)$. The case $n = 8$ is handled by Example 2.2. So let $12k + 8 \geq 20$. It is well kown that there is a PBD with block sizes 3 and 4 for every $n \equiv 4$ (mod 6) [13]. Let (X, B) be such a PBD, $|X| \equiv 4$ (mod 6), and proceed exactly as in the cases for $n \equiv 2$ or 6 (mod 12), using Example 2.2 as well as Example 2.1.

Lemma 3.1. *There exists a max packing of K_n with 6-cycles that can be squashed into a max packing of K_n with triples for all $n \equiv 2, 6, 8 \pmod{12} \geq 6$.* □

3.2 $n \equiv 0$ (mod 12)

We begin with an example.

Example 3.2. **($n = 12$)**

Let $X = \{\infty_1, \infty_2\} \cup Z_{10}$ and define a collection of 6-cycles C as follows:

$(0, \infty_1, 2, 1, 3, 6)$, $(2, \infty_2, 4, 3, 5, 8)$, $(4, \infty_1, 6, 5, 7, 0)$, $(6, \infty_2, 8, 7, 9, 2)$
$(8, \infty_1, 9, 0, 1, 4)$, $(1, \infty_1, 3, 0, 2, 7)$, $(3, \infty_2, 5, 2, 4, 9)$, $(5, \infty_1, 7, 4, 6, 1)$,
$(7, \infty_2, 9, 6, 8, 3)$, $(0, \infty_2, 1, 8, 9, 5)$,

with leave $L = \{(0, 8), (1, 9), (2, 3), (4, 5), (6, 7), (\infty_1, \infty_2)\}$. Then (X, C) is a max packing of K_{12} with 6-cycles and can be squashed into a max packing of K_{12} with triples with leave L.

We will need two constructions for $12k \geq 24$: one when k is even and one when k is odd.

$12k$, k even Write $12k = 4(3k)$ and let (P, G, B) be a GDD$(3k, \{2\}, 3)$, set $X = P \times \{1, 2, 3, 4\}$ and define a collection of 6-cycles C as follows:

(i) For each group $g \in G$ place Example 2.2 on $g \times \{1, 2, 3, 4\}$ with leave $L_g = \{g \times \{1\}, g \times \{2\}, g \times \{3\}, g \times \{4\}\}$ and place these 6-cycles in C.

(ii) For each triple $t = \{a, b, c\} \in B$ place a copy of Example 2.3 on $K_{4,4,4}$ with parts $\{a\} \times \{1, 2, 3, 4\}$, $\{b\} \times \{1, 2, 3, 4\}$, and $\{c\} \times \{1, 2, 3, 4\}$ and place these 6-cycles in C.

Then (X, C) is a max packing of K_{12k} with 6-cycles with leave $L = \{g \times \{1\}, g \times \{2\}, g \times \{3\}, g \times \{4\} \mid g \in G\}$. It is straightforward to see that the 6-cycles in (i) and (ii) can be squashed into a max packing of K_{12k} with triples with leave L.

$12k$, k odd Write $12k = 4(3k)$. Since k is odd, $3k$ is the order of a Kirkman triple system (P, T). Let $X = P \times \{1, 2, 3, 4\}$, π a parallel class in T, and define a collection of 6-cycles C as follows:

(i) For each triple $t = \{a, b, c\} \in \pi$, place a copy of Example 3.2 on $\{a, b, c\} \times \{1, 2, 3, 4\}$ with leave L_t and place these 6-cycles in C.

(ii) For each triple $\{a, b, c\} \in T \setminus \pi$, place a copy of Example 2.3 on $K_{4,4,4}$ with parts $\{a\} \times \{1, 2, 3, 4\}$, $\{b\} \times \{1, 2, 3, 4\}$, and $\{c\} \times \{1, 2, 3, 4\}$, and place these 6-cycles in C.

Then (X, C) is a max packing of K_{12k} with 6-cycles with leave $L = \{L_t | \, t \in \pi\}$. Squashing these 6-cycles produces a max packing of K_{12k} with triples with leave L.

Lemma 3.3. *There exists a max packing of K_n with 6-cycles that can be squashed into a max packing of K_n with triples for all $n \equiv 0 \pmod{12}$.* □

3.3 $n \equiv 11 \pmod{12}$

We begin with an example.

Example 3.4. (n = 11)

Let $X = Z_9 \cup \{\infty_1, \infty_2\}$ and define a collection of 6-cycles C as follows:

$(4, 8, \infty_2, 7, \infty_1, 0)$, $(5, 0, \infty_2, 4, \infty_1, 1)$, $(6, 1, \infty_2, 5, \infty_1, 2)$, $(7, 2, \infty_2, 6, \infty_1, 3)$
$(3, 2, 5, 7, 0, 1)$, $(7, 4, 5, 3, 0, 6)$, $(4, 3, 6, 8, 1, 2)$, $(8, 5, 6, 4, 1, 7)$

with leave $L = \{(\infty_1, \infty_2, 3, 8), (0, 2, 8)\}$. Then (X, C) is a max packing of K_{11} with 6-cycles with leave L. Squashing these 6-cycles and adding (0, 2, 8) from the leave L gives a max packing of K_{11} with triples with leave the 4-cycle $(\infty_1, \infty_2, 3, 8)$.

We can now give a general construction for 11 (mod 12) ≥ 23.

$\underline{12k + 11 \geq 23}$ Write $12k + 11 = 2(6k + 4) + 3$. Let (P, G, B) be a GDD$(6k + 4, \{4^*, 2\}, 3)$ [13], set $X = \{\infty_1, \infty_2, \infty_3\} \cup (P \times \{1, 2\})$, and define a collection of 6-cycles C as follows:

(i) Let b^* be the unique group of size 4 and define a copy of Example 3.4 on $\{\infty_1, \infty_2, \infty_3\} \cup (b^* \times \{1, 2\})$ with leave $L = \{(\infty_1, \infty_2, \infty_3), (x, y, z, w)\}$, where $\{x, y, z, w\} \subseteq b^* \times \{1, 2\}$ and place these 6-cycles in C.

(ii) For each group $g \in G$ of size 2, define a copy of a max packing of K_7 with 6-cycles, with vertex set $\{\infty_1, \infty_2, \infty_3\} \cup (b \times \{1, 2\})$, that can be squashed into 6-triples with leave $(\infty_1, \infty_2, \infty_3)$ [11]. Add these 6-cycles to C.

(iii) For each triple $t = \{a, b, c\} \in B$, place a copy of Example 2.1 on $t \times \{1, 2\}$ with leave $\{a\} \times \{1, 2\}$, $\{b\} \times \{1, 2\}$, and $\{c\} \times \{1, 2\}$ and place these 6-cycles in C.

Then (X, C) is a max packing of K_{12k+11} with 6-cycles with leave L in (i). If we squash these 6-cycles and add the triple $(\infty_1, \infty_2, \infty_3)$ from the leave L in (i) we have a max packing of K_{12k+11} with triples with leave (x, y, z, w) in (i).

Lemma 3.5. *There exists a max packing of K_n with 6-cycles that can be squashed into a max packing of K_n with triples for all $n \equiv 11 \pmod{12}$.* □

3.4 $n \equiv 4$ or 10 (mod 12)

The following three examples are necessary for the constructions in this section.

Example 3.6. ($n = 10$)

Let $X = \{\infty\} \cup (Z_3 \times Z_3)$ and define a collection of 6-cycles C as follows: $(0_1, 1_1, 0_0, 1_2, 0_2, \infty)$, $(1_2, 2_1, 0_0, 0_1, 0_2, 1_0)$, $(1_1, 2_1, 1_0, 2_2, 1_2, \infty)$, $(2_2, 0_1, 1_0, 1_1, 1_2, 2_0)$, $(2_1, 0_1, 2_0, 0_2, 2_2, \infty)$, $(0_2, 1_1, 2_0, 2_1, 2_2, 0_0)$ with leave $L = \{\{\infty, 2_0, 1_0, 0_0\}, (0_1, 1_2), (1_1, 2_2), (2_1, 0_2)\}$. (We remark that $\{\infty, 2_0, 1_0, 0_0\}$ is a copy of K_4 and not a 4-cycle.) Then (X, C) is a max packing of K_{10} with 6-cycles with leave L. If we squash these 6-cycles and remove a triple from $\{\infty, 2_0, 1_0, 0_0\}$, the result is a max packing of K_{10} with triples with leave the tripole $K_{1,3} \cup \{(0_1, 1_2), (1_1, 2_2), (2_1, 0_2)\}$.

Example 3.7. ($n = 16$)

Let $X = \{\infty_1, \infty_2, \infty_3, \infty_4\} \cup \{i_j \mid i \in Z_6, j \in \{0, 1\}\}$. Further, for each $i \in Z_6$, define $\alpha(i) = \infty_1$ if i is odd and ∞_2 if i is even. For each $i \in Z_6$ define a collection of 6-cycles C as follows: $(i_1, i_0, (4 + i)_1, (2 + i)_1, (1 + i)_1, \alpha(i))$, $(i_0, (1 + i)_1, (4 + i)_0, (2 + i)_0, (1 + i)_0, \alpha(i))$, and $(i_0, (2 + i)_1, \infty_3, (1 + i)_0, \infty_4, (5 + i)_1)$ with leave $L = \{\infty_1, \infty_2, \infty_3, \infty_4\} \cup \{(i_j, (3 + i)_j) \mid i \in \{0, 1, 2\}, j \in \{0, 1\}\}$. (Once again we remark that $\{\infty_1, \infty_2, \infty_3, \infty_4\}$ is a copy of K_4.) Then (X, C) is a max packing of K_{16} with 6-cycles with leave L. If we squash these 6-cycles and remove a triple from $\{\infty_1, \infty_2, \infty_3, \infty_4\}$, the result is a max packing of K_{16} with triples, with leave the tripole $K_{1,3} \cup \{(i_j, (3 + i)_j) \mid i \in \{0, 1, 2\}, j \in \{0, 1\}\}$.

Example 3.8. ($n = 28$)

Let $X = \{\infty_1, \infty_2, \infty_3, \infty_4\} \cup (Z_{12} \times \{1,2\})$, and let (P, G, B) be a GDD$(12, \{3\}, 4)$ (equivalent to a pair of orthogonal quasigroups of order 3) and define a collection of 6-cycles C as follows:

(i) For each group $g \in G$, place a copy of Example 3.6 on $\{\infty_1, \infty_2, \infty_3, \infty_4\} \cup (g \times \{1,2\})$ with leave $\{\infty_1, \infty_2, \infty_3, \infty_4\} \cup \{(x_1, x_2) \mid x \in g\}$ (with K_4 based on $\{\infty_1, \infty_2, \infty_3, \infty_4\}$).

(ii) For each block $b \in B$ place a copy of Example 2.2 on $b \times \{1,2\}$ with leave $\{\{x_1, x_2\} \mid x \in b\}$.

Then (X, C) is a max packing of K_{28} with 6-cycles with the leave in (i). Now squashing the 6-cycles in (i) and (ii) and removing a triple from K_4 gives a max packing of K_{28} with triples with leave a tripole.

We can now go to the general constructions for $n \equiv 10 \pmod{12}$, $n \geq 22$ and $n \equiv 4 \pmod{12}$, $n \geq 40$.

<u>$n \equiv 10 \pmod{12}$, $n \geq 22$</u> Write $12k + 10 = 2(6k + 5)$ and let (P, B) be a PBD$(6k + 5, \{5^*, 3\})$ [13]. Set $X = P \times \{1,2\}$ and define a collection C of 6-cycles as follows:

(i) Let b^* be the unique block of size 5 and define a copy of Example 3.6 on $b^* \times \{1,2\}$ and place these 6-cycles in C. (The leave is $K_4 \cup \{$1-factor$\}$.)

(ii) For each triple $t = \{a, b, c\} \in B$, define a copy of Example 2.1 on $t \times \{1,2\}$ with leave $\{(a_1, a_2), (b_1, b_2), (c_1, c_2)\}$ and place these 6-cycles in C.

Then (X, C) is a max packing of K_{12k+10} with 6-cycles with leave $K_4 \cup \{$1-factor$\}$. Squashing the 6-cycles in C and removing a triple from the leave in (i) produces a max packing of K_{12k+10} with triples with leave a tripole.

<u>$n \equiv 4 \pmod{12}$, $n \geq 40$</u> Write $12k+4 = 4+2(6k)$ and let (P, G, B) be a GDD$(6k, \{6\}, 3)$ [13]. Set $X = \{\infty_1, \infty_2, \infty_3, \infty_4\} \cup (P \times \{1,2\})$ and define a collection of 6-cycles as follows:

(i) For each group $g \in G$ define a copy of Example 3.7 on$\{\infty_1, \infty_2, \infty_3, \infty_4\} \cup (g \times \{1,2\})$ with leave $K_4 \cup \{(x_1, x_2) \mid x \in G\}$ (K_4 is based on $\{\infty_1, \infty_2, \infty_3, \infty_4\}$).

(ii) For each triple $t = \{a, b, c\} \in B$, define a copy of Example 2.1 on $t \times \{1,2\}$ with leave $\{(a_1, a_2), (b_1, b_2), (c_1, c_2)\}$ and place these 6-cycles in C.

Then (X, C) is a max packing of K_{12k+4} with 6-cycles. Squashing the 6-cycles in C and removing a triple from the leave K_4 in (i) produces a max packing of K_{12k+4} with triples with leave a tripole.

Lemma 3.9. *There exists a max packing of K_n with 6-cycles that can be squashed into a max packing of K_n with triples for all $n \equiv 4$ or $10 \pmod{12} \geq 6$.* □

3.5 $n \equiv 5 \pmod{12}$

This case requires three examples.

Example 3.10. $(n = 17)$

Let $X = \{\infty_1, \infty_2\} \cup Z_{15}$ and define a collection of 6-cycles C as follows: $\{(0,9,4,5,1,3)+i \mid i \in Z_{15}\} \cup \{(7,14,\infty_2,13,\infty_1,0), (8,0,\infty_2,7,\infty_1,1), (9,1,\infty_2,8,\infty_1,2), (10,2,\infty_2,9,\infty_1,3), (11,3,\infty_2,10,\infty_1,4), (12,4,\infty_2,11,\infty_1,5), (13,5,\infty_2,12,\infty_1,6)\}$ with leave the 4-cycle $(\infty_1,\infty_2,6,14)$. Squashing all of the 6-cycles in C produces a max packing of K_{17} with triples with leave the 4-cycle $(\infty_1,\infty_2,6,14)$.

Example 3.11. $(n = 29)$

Let $X = \{\infty_1, \infty_2\} \cup Z_{27}$ and define a collection of 6-cycles C as follows: $\{(0,3,1,5,4,9)+i, (0,16,10,17,7,15)+i \mid i \in Z_{27}\} \cup \{(14,0,\infty_2,13,\infty_1,1)+j \in \{0,1,2,...,11\}, (13,26,\infty_2,25,\infty_1,0)\}$ with leave the 4- cycle $(\infty_1,\infty_2,12,26)$. Squashing the 6-cycles in C gives a max packing of K_{29} with triples with leave the 4-cycle $(\infty_1,\infty_2,12,26)$.

Example 3.12. $(n = 53)$

Let $X = \{\infty_1, \infty_2\} \cup Z_{51}$ and define a collection of 6-cycles C as follows: $\{(0,3,1,5,4,21)+i, (0,19,9,20,11,23)+i, (0,24,8,15,7,22)+i, (0,20,6,11,5,18)+i \mid i \in Z_{51}\} \cup \{(26,0,\infty_2,25,\infty_1,1)+j \mid j \in \{0,1,2,\ldots,23\}\} \cup \{(25,50,\infty_2,49,\infty_1,0)\}$ with leave the 4-cycle $(\infty_1,\infty_2,24,50)$. Squashing these 6-cycles gives a max packing of K_{53} with triples with leave the 4-cycle $(\infty_1,\infty_2,24,50)$.

We can now give two general constructions to finish off the case $n \equiv 5$ (mod 12).

<u>$12k+5$, k odd</u> Write $12k+5 = 1+4(3k+1)$. Since k is odd, $1+4(3k+1) = 1+4(6t+4)$. Let (P,G,B) be a GDD$(6t,\{4^*,2\},3)$, set $X = \{\infty\} \cup (P \times \{1,2,3,4\})$ and define a collection of 6-cycles as follows:

(i) For the unique group b^* of size 4, define a copy of Example 3.10 on $\{\infty\} \cup (b^* \times \{1,2,3,4\})$ (the leave is a 4-cycle) and place these 6-cycles in C.

(ii) For each group g of size 2, define a copy of Example 1.1 on $\{\infty\} \cup (g \times \{1,2\})$ and place these 6-cycles in C. (There is no leave.)

(iii) For each triple $\{a,b,c\} \in B$, place a copy of Example 2.3 on $\{a,b,c\} \times \{1,2,3,4\}$ with parts $\{a\} \times \{1,2,3,4\}$, $\{b\} \times \{1,2,3,4\}$, $\{c\} \times \{1,2,3,4\}$, and place these 6-cycles in C. (There is no leave.)

Then (X,C) is a max packing of K_{12k+5} with 6-cycles with leave a 4-cycle. If we squash the 6-cycles in (i), (ii) and (iii), we have a max packing of K_{12k+5} with triples with leave a 4-cycle.

<u>$12k+5$, k even</u> Write $12k+5 = 1+4(3k+1)$. Since k is even, $1+4(3k+1) = 1+4(6t+1)$. Since $12k+5 \geq 77$, $6t+1 \geq 19$, and there exists a GDD$(6t+1,\{7^*,3\},3)$ [11] (P,G,B). Define a collection C of 6-cycles on $X = \{\infty\} \cup (P \times \{1,2,3,4\})$ as follows:

(i) For the unique group b^* of size 7, define a copy of Example 3.11 on $\{\infty\} \cup (b \times \{1,2,3,4\})$ (with leave a 4-cycle) and place these 6-cycles in C.

(ii) For each group g of size 3, place a copy of a 6-cycle system of order 13 which can be squashed into a triple system [11] (no leave) on $\{\infty\} \cup (g \times \{1,2,3,4\})$ and place these 6-cycles in C.

(iii) For each triple $\{a,b,c\} \in B$, place a copy of Example 2.3 on $\{a,b,c\} \times \{1,2,3,4\}$ with parts $\{a\} \times \{1,2,3,4\}$, $\{b\} \times \{1,2,3,4\}$, $\{c\} \times \{1,2,3,4\}$, and place these 6-cycles in C. (There is no leave.)

Then (X, C) is a max packing of K_{12k+5} with 6-cycles with leave the 4-cycle in (i). Squashing the 6-cycles in (i), (ii) and (iii) produces a max packing of K_{12k+5} with triples with leave the 4-cycle in (i).

Lemma 3.13. *There exists a max packing of K_n with 6-cycles that can be squashed into a max packing of K_n with triples for all $n \equiv 5 \pmod{12} \geq 17$.* □

4 Main result and further developments

Putting together the results in Section 3 gives the following theorem.

Theorem 4.1. *For each $n \equiv 0, 2, 4, 5, 6, 8, 10, 11 \pmod{12}$, $n \geq 6$, there exists a max packing of K_n with 6-cycles that can be squashed into a max packing of K_n with triples. There are no exceptions.* □

Since a complete solution is also a max packing, we can combine Theorem 1.3 and Theorem 8.1 into the following corollary (giving a complete solution to the squashing of max packings of 6-cycles into max packings with triples).

Corollary 4.2. *For each $n \geq 6$, there is a max packing of K_n with 6-cycles that can be squashed into a max packing of K_n with triples.* □

In this paper we give a complete solution to the problem of squashing maximum packings of K_n with 6-cycles into maximum packings of K_n with triples. An open problem is to solve the general case, i.e. squashing a maximum packing of K_n with $2m$-cycles into a maximum packing of K_n with m-cycles; the case $m = 4$ is completely solved in [10] .

References

[1] L. Berardi, M. Gionfriddo and R. Rota, Perfect octagon quadrangle systems - II, *Discrete Math.* **312**(3) (2012), 614–620.

[2] Y. Chang, T. Feng, G. Lo Faro and A. Tripodi, The existence spectrum of Meta$(K_4^{(3)} > K_4^{(3)} - e)$, *Science China Mathematics* **11** (53) (2010), 2865–2876.

[3] Y. Chang, T. Feng, G. Lo Faro and A. Tripodi, The triangle intersection numbers of a pair of disjoint S(2, 4, v)s, *Discrete Math.* **310**(21) (2010), 3007–3017.

[4] Y. Chang, T. Feng, G. Lo Faro and A. Tripodi, The fine triangle intersection for $(K_4 - e)$-designs, *Discrete Math.* **311**(21) (2011), 2442–2462.

[5] Y. Chang, T. Feng, G. Lo Faro and A. Tripodi, The fine triangle intersection for kite systems, *Discrete Math.* **312** (3) (2012), 545–553.

[6] Y. Chang, G. Lo Faro and A. Tripodi, Tight blocking sets in some maximum packings of λK_n, *Discrete Math.* **308** (2008), 427–438.

[7] Y. Chang, G. Lo Faro and A. Tripodi, Determining the spectrum of $Meta(K_4 + e > K_4, \lambda)$ for any λ, *Discrete Math.* **308** (2008), 439–456.

[8] J. A. Kennedy, Maximum packings of K_n with hexagons, *Australas. J. Combin.* **7** (1993), 101–110.

[9] T. P. Kirkman, On a problem in combinations, *Cambridge and Dublin Math. J.* **2** (1847), 191–204.

[10] C. C. Lindner, G. Lo Faro, M. Meszka and A. Tripodi, Squashing maximum packings of K_n with 8-cycles into maximum packings of K_n with 4-cycles, *Filomat* **28:4** (2014), 887–896.

[11] C. C. Lindner, M. Meszka and A. Rosa, From squashed 6-cycles to Steiner triple systems, *J. Combin. Des.* **22** (2014), 189–195.

[12] C. C. Lindner, K. T. Phelps and C. A. Rodger, The spectrum for 2-perfect 6-cycle systems, *J. Combin. Theory Ser. A* **57** (1991), 76–85.

[13] C. C. Lindner and C. A. Rodger, *Design Theory*, second ed., CRC Press, 2009.

[14] A. Rosa, On cycle decompositions of the complete graph into $(4m+2)$-gons, *Mat-fyz Časopis* **16** (1966), 349–435.

[15] A. Rosa and C. Huang, Another class of balanced graph designs: balanced circuit designs, *Discrete Math.* **12** (1975), 269–293.

ARS MATHEMATICA CONTEMPORANEA

Also available at http://amc-journal.eu
ISSN 1855-3966 (printed edn.), ISSN 1855-3974 (electronic edn.)
ARS MATHEMATICA CONTEMPORANEA 10 (2016) 31–44

Multicoloring of cannonball graphs

Petra Šparl *

University of Maribor, Faculty of Organizational Sciences, Kidričeva 55a, SI-4000 Kranj, Slovenia
and
University of Primorska, IAM, Muzejski trg 2, 6000 Koper, Slovenia

Rafał Witkowski †

Adam Mickiewicz University, Faculty of Mathematics and Computer Science, ul. Umultowska 87, 61-614 Poznań, Poland

Janez Žerovnik ‡

University of Ljubljana, Faculty of Mechanical Engineering, Aškerčeva 6, SI-1000 Ljubljana, Slovenia
and
IMFM, Jadranska 19, Ljubljana, Slovenia

Received 2 September 2013, accepted 12 March 2014, published online 19 March 2015

Abstract

The frequency allocation problem that appeared in the design of cellular telephone networks can be regarded as a multicoloring problem on a weighted hexagonal graph, which opened some still interesting mathematical problems. We generalize the multicoloring problem into higher dimension and present the first approximation algorithms for multicoloring of the so called cannonball graphs.

Keywords: Graph coloring, approximation algorithm, frequency planning, cellular networks.

Math. Subj. Class.: 05C15, 05C85, 68W25, 68R10

* Supported in part by the Slovenian Research Agency (research project J1-5433).
† This work was supported by grant N206 1842 33 for years 2011-2014
‡ Supported in part by ARRS, the research agency of Slovenia.

E-mail addresses: petra.sparl@fov.uni-mb.si (Petra Šparl), rmiw@amu.edu.pl (Rafał Witkowski), janez.zerovnik@fs.uni-lj.si, janez.zerovnik@imfm.si (Janez Žerovnik)

1 Introduction

A fundamental problem that appears in the design of cellular networks is to assign sets of frequencies to transmitters in order to avoid the unacceptable interferences. The number of frequencies demanded at a transmitter may vary between transmitters. The problem appeared in the sixties and was soon related to the graph multicoloring problem (the formal definition is given on page 33), see the early survey [5]. It received an enormous attention in the nineties and is still of considerable interest (see [1] and the references there). Besides the mobile telephony there are several applications of frequency assignment including radio and television broadcasting, military applications, satellite communication and wireless LAN [1]. A sizable part of theoretical studies is concentrated on the simplified model when the underlying graph which has to be multicolored is a subgraph of triangular grid (see [12, 13, 5]). This is a natural choice because it is well known that hexagonal cells provide a coverage with the optimal ratio of the distance between centers compared to the area covered by each cell. Such graphs are called hexagonal graphs [18, 19, 21]. Indeed, the model is a reasonable approximation for the rural cellular networks where the underlying graph is often nearly planar, and a popular example are the sets of benchmark problems based on the real cellular network around Philadelphia [2] (see the FAP website [29]).

Although the multicoloring (the formal definition is given on page 33) of hexagonal graphs seems to be a very simplified optimization problem, some interesting mathematical questions were asked at the time that are still open. An example is the conjecture of McDiarmid and Reed saying that the multichromatic number (the formal definition is given on page 33) of any hexagonal graph G is between $\omega(G)$ and $9\omega(G)/8$, where $\omega(G)$ is the weighted clique number [12]. On the other hand, the hexagonal graph model is known to be practically useless in urban areas, where high concrete buildings on the one hand prevent propagation of the radio signals and on the other hand allow very high concentration of users. Loosely speaking, a three dimensional model may be needed in contrast to the hexagonal graphs that are a good model for two dimensional networks. In this paper we discuss a generalization of the multicoloring problem on hexagonal graphs from the planar case to three dimensions. It is well known that hexagonal cells of the same size with centers positioned in the triangular grid provide an optimal coverage of the plane. Optimality here means the best ratio between the diameter and the area covered by the cell. The situation is much more interesting in three dimensions. Obviously, optimal cells would be nearly balls, and the question is how to position the centers of the balls to achieve the optimal diameter to volume ratio. The famous Kepler conjecture was a longstanding conjecture about the ball packing in three-dimensional Euclidean space. It says that no arrangement of equally sized balls filling space has greater average density than that of the cubic close packing (face-centered cubic) and the hexagonal close packing arrangements. The density of these arrangements is slightly greater than 74%. It may be interesting to note that the solution of Kepler's conjecture is included as a part of the 18th problem in the famous Hilbert's problem list back in 1900 [22]. Recently Thomas Hales, following an approach suggested by Fejes Tóth, published a proof of the Kepler conjecture. For more details, see [6, 7].

Given an optimal arrangement of balls, we define a graph by taking the balls (or centers of balls) as vertices and connect each pair of touching balls with an edge. Nonnegative demands are assigned to each vertex and we are interested in multicoloring of the graph induced on vertices of positive demand. Loosely speaking, we generalize the problem of multicoloring of hexagonal graphs from two dimensions to three dimensions. The question "What ratio $\chi(G)/\omega(G)$ can be obtained by a generalization of 2-dimensional algorithm to 3-dimensional algorithm" has been asked at the Oberwolfach seminar Algorithmische Graphentheorie [28] and we are not aware of any result since then.

More formally, we are interested in multicoloring of weighted graphs $G = (V(G), E(G), d)$, where $V = V(G)$ is the set of vertices, $E = E(G)$ is the set of edges, and d assigns a positive integer $d(v)$ to vertex $v \in V$. $d(v)$ is the *weight* of a vertex, here also called *demand*. Adjacent vertices are called *neighbors*. The *degree* of a vertex, $deg_G(v) = deg(v)$ is the number of neighbors of v. A *proper multicoloring* of G is a mapping f from $V(G)$ to subsets of integers such that $|f(v)| \geq d(v)$ for any vertex $v \in V(G)$ and $f(v) \cap f(u) = \emptyset$ for any pair of adjacent vertices u and v in the graph G. The minimum number of colors needed for a proper multicoloring of G, $\chi_m(G)$, is called the *multichromatic number*. Another invariant of interest in this context is the *(weighted) clique number*, $\omega(G)$, defined as follows: The weight of a clique of G is the sum of demands on its vertices and $\omega(G)$ is the maximal clique weight on G. Clearly, $\chi_m(G) \geq \omega(G)$. *Hexagonal graph* is the graph induced on vertices of triangular grid of positive demand. Or, in other words, cells of hexagonal grid are assigned non-negative integer demands, and the graph is composed by taking cells with positive demand as vertices and two vertices whose hexagons share an edge are regarded to be adjacent. In the 3-dimensional case we will consider optimal arrangements of balls, and define a graph by taking balls (with positive demand) as vertices, and connect touching balls by edges. We call these graphs the *cannonball graphs* as Keplers motivation for studying the arrangements of balls was optimal arrangement of cannonballs. McDiarmid and Reed proved in [12] that multicoloring of hexagonal graphs is NP-complete. In the last decade there were several results on upper bounds for the multichromatic number in terms of weighted clique number for hexagonal graphs, some of which also provide approximation algorithms that are fully distributed and run in constant time [8, 9, 10, 12, 13, 14, 16, 18, 19, 20, 15, 21, 23, 24, 25, 27]. The best known approximation ratios are $\chi_m(G) \leq (4/3)\omega(G) + O(1)$ in general [12, 14, 18] and $\chi_m(G) \leq (7/6)\omega(G)+O(1)$ for triangle free hexagonal graphs [8, 15, 16]. The conjecture of McDiarmid and Reed: $\chi_m(G) \leq (9/8)\omega(G) + O(1)$ remains an open problem [12].

Since multicoloring of cannonball graphs is an extension of multicoloring of hexagonal graphs, it is NP-complete. No approximation algorithm and no upper bound was previously known for the multichromatic number of cannonball graphs. Here we give two upper bounds, where the first is easily implied by known results for hexagonal graphs (because a layer in a cannonball graph is a hexagonal graph) and the second is an improvement of the first upper bound using some structural properties of the cannonball graphs. In both cases, the constructions are given, thus providing polynomial-time approximation algorithms. The main result of this paper that gives the first answer to a problem asked in [28] is

Theorem 1.1. *There is a polynomial-time approximation algorithm for multicoloring cannonball graphs which uses at most* $\frac{11}{6}\omega(G) + \frac{25}{6}$ *colors.*

The paper is organized as follows. In the next section we formally define some basic terminology. In Section 3, we present an overview of the algorithm, while in Section 4 we provide a proof of Theorem 1.1. In the last Section we give some ideas for further work.

2 Hexagonal and cannonball graphs

First we formally define hexagonal and cannonball graphs. Recall the formal definition of hexagonal graphs: the position of each vertex is an integer linear combination $x\vec{p} +$

$y\vec{q}$ of two vectors $\vec{p} = (1, 0)$ and $\vec{q} = (\frac{1}{2}, \frac{\sqrt{3}}{2})$ and the vertices of the triangular grid are identified with pairs (x, y) of integers. Put an edge connecting two vertices if the points representing the vertices are at Euclidean distance one in the triangular grid (in other words, when the corresponding hexagonal cells are adjacent). To construct a hexagonal graph G, positive weights are assigned to a finite subset of points in the grid and G is the subgraph induced on $V(G)$, the set of grid vertices with positive weights. Cannonball graphs are constructed in a similar way. However, we have many possibilities already when constructing the underlying grid, which, loosely speaking, consists of tetrahedrons and will be called *tetrahedron grid* $T \in \mathcal{G}$, where $\mathcal{G}$ is an infinite family of such grids. We will construct a cannonball graph starting from a fixed grid, which in turn however can be one of many possible grids that arrise from optimal ball packings.

Optimal arrangement of balls in one layer is to put the centers of the balls in the points of triangular grid. Then, there are exactly two possibilities to put a second layer on the top of the first layer. These two arrangements are obviously symmetric, however, when choosing a position for the third layer, there are two possibilities that give rise to different arrangements. We will call them *layer-arrangement (a)* and *layer-arrangement (b)*, respectively (see figure 1).

Consequently, we have an infinite number of tetrahedron grids, that all came from the optimal ball arrangements. One of the arrangements, called the cubic close packing (see case (a) of figure 1), can be described nicely by introducing a third vector $\vec{r} = (\frac{1}{2}, \frac{\sqrt{3}}{6}, \frac{\sqrt{6}}{3})$ in addition to $\vec{p} = (1, 0, 0)$ and $\vec{q} = (\frac{1}{2}, \frac{\sqrt{3}}{2}, 0)$. Now the position of each vertex is an integer linear combination $x\vec{p} + y\vec{q} + z\vec{r}$ and the vertices of the tetrahedron grid may be identified with triplets (x, y, z) of integers. Given the vertex v, we will refer to its coordinates as $x(v)$, $y(v)$ and $z(v)$, or shortly x, y, and z, when there is no confusion possible. For other arrangements there is no such easy extension of the notation from hexagonal graphs. A cannonball graph G is obtained by assigning integer weights to the points of the tetrahedron grid T, taking as $V(G)$ the vertices in the grid with positive weights, and introducing edges between vertices at Euclidean distance one (in other words, connecting the touching balls). The cannonball graphs based on the cubic close packing will be called *regular cannonball graphs*. Clearly, from the construction it follows that any layer of an arbitrary cannonball graph is a hexagonal graph (maybe not connected).

Formally, a *cannonball graph* is a graph induced on vertices of positive weight.

There is a natural basic 4-coloring of (unweighted) cannonball graph. Start with any layer and call it the base layer. Introduce coordinates $(x, y, 0)$ in this layer and define the base coloring by the formula

$$bc(v) = x \bmod 2 + 2(y \bmod 2). \tag{2.1}$$

Colors of vertices of the next layers are then determined exactly as follows. It is obvious that whenever we store a new layer above (or under) the previous one with fixed coloring, we know that each ball from the new layer is connected to exactly three balls from the previous layer, and all of those balls have different colors. Thus there is exactly one extension of the four coloring to the next layer (see figures 1 and 2, where 4-coloring, using colors $0, 1, 2, 3$, is presented).

It is easy to see that this rule, starting from (2.1), gives a proper coloring of the next layers. In regular cannonball graphs this coloring can be given by closed expression in the following way:

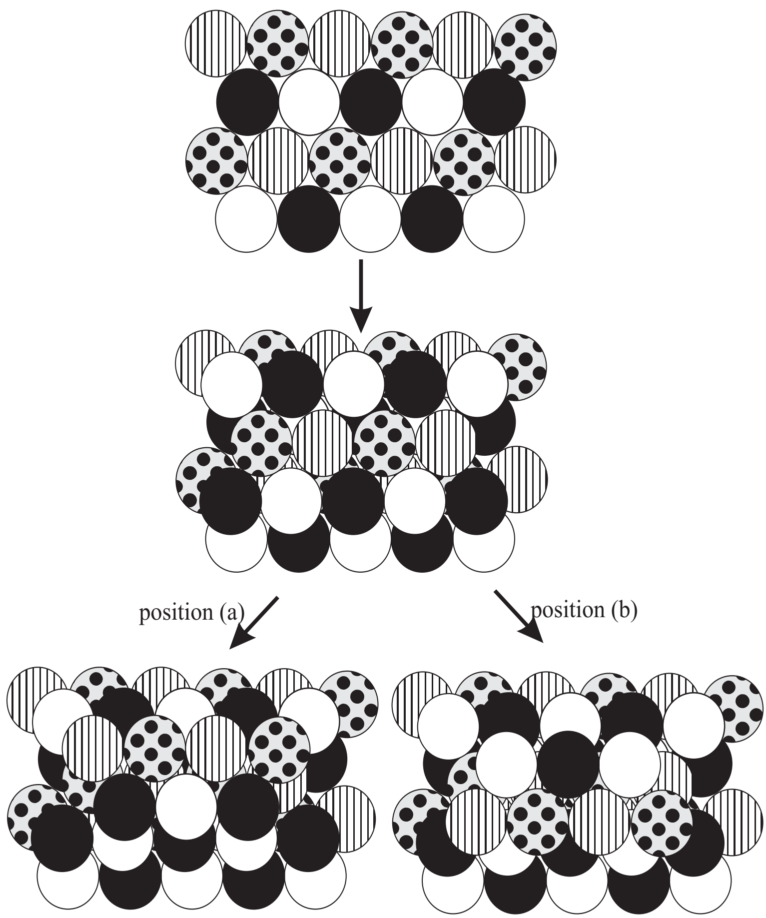


Figure 1: Two different arrangements of the third layer.

$$bc(v) = ((z+1) \bmod 2)(x \bmod 2 + 2(y \bmod 2))+ \\ + (z \bmod 2)((x+1) \bmod 2 + 2((y+1) \bmod 2)). \quad (2.2)$$

From the construction of cannonball graphs it is clear that each vertex has (at most) 6 neighbors in its layer, and in addition (at most) three neighbors in each of the neighboring layers. The degree of a vertex in cannonball graph is hence at most 12 (see figure 2).

The cliques in the cannonball graphs can have at most four vertices. The *(weighted) clique number*, $\omega(G)$, is the maximal clique weight on G, where the weight of a clique is the sum of weights on its vertices. As cliques in cannonball graphs can have at most four vertices, the weighted clique number is the maximum weight over weights of all tetrahedrons, triangles, edges and weights of isolated vertices. Therefore, we can define invariants $\omega_i(G)$ which denote the maximum weight of a clique of size at most i on G. In fact, we can regard the clique numbers as based on the complete subgraphs of the grid graph because

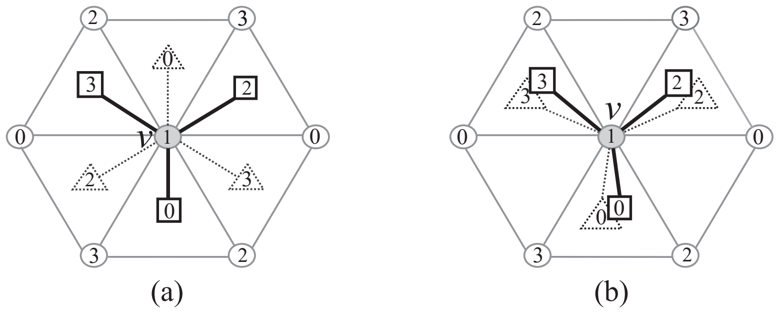


Figure 2: All twelve possible neighbors of vertex $v \in G$ for arrangements (a) and (b). Circles and gray lines represent the middle layer containing v, squares and thick lines represent the upper layer, dashed triangles and dashed lines represent the lower layer.

the vertices of weight 0 clearly do not contribute to the clique weights. For example, $\omega_2(G)$ is the maximal weight over all edges and isolated vertices. Clearly, for cannonball graphs we have

$$\omega_1(G) \leq \omega_2(G) \leq \omega_3(G) \leq \omega_4(G) = \omega(G).$$

An induced subgraph of the cannonball graph without a 3-clique will be called a *triangle-free cannonball graph.*

In the algorithm we will consider some subgraphs of the cannonball graph, in particular, it may be useful to have 3-colorable subgraphs.

For 3-colorable graphs, there is a simple multicoloring algorithm that uses at most $\lceil \frac{3}{2}\omega(G) \rceil$ colors.

Lemma 2.1. *Every 3-colorable graph G can be multicolored using at most $\lceil \frac{3}{2}\omega(G) \rceil$ colors.*

We will prove Lemma 2.1 using the following procedure.

Procedure 2.2. Let G be an arbitrary 3-colorable graph with coloring $c : V(G) \to \{1, 2, 3\}$. Define $K = \left\lfloor \frac{\omega(G)}{2} \right\rfloor$. Assign $\min\{d(v), K\}$ colors to every vertex. More precisely, a vertex of color $c(v)$ receives colors from the set $\{c(v), c(v)+3, \ldots, c(v)+3(K-1)\}$.
a) For the case $\omega = 2k + 1$ do the following:
If $d(v) \geq K + 1$ then assign the additional color 0 to v and if $d(v) \geq K + 2$ assign to v the highest $N = d(v) - K - 1$ colors from one of its neighbors' palettes. More precisely, take $b \in \{1, 2, 3\} \backslash c(v)$ and use the colors $\{b + 3K, b + 3(K - 1), \ldots, b + 3(K - N + 1)\}$.
b) For the case $\omega = 2k$ do the following:
If $d(v) \geq K + 1$ assign to v the highest $N = d(v) - K$ colors from one of its neighbors' palettes. More precisely, take $b \in \{1, 2, 3\} \setminus c(v)$ and use colors $\{b + 3K, b + 3(K - 1), \ldots, b + 3(K - N + 1)\}$.

Proof of Lemma 2.1. Note that if $d(v) > K$ then for any neighbor u of v we have $d(u) \leq K$. Furthermore, $d(u) + N = d(u) + d(v) - 1 - K \leq \omega(G) - 1 - K \leq K$, and there is no conflict possible.

a) Suppose $\omega = 2k+1$. Then in Procedure 2.2 all together we need at most $3\left\lfloor \frac{\omega(G)}{2} \right\rfloor + 1 < \left\lceil \frac{3}{2}\omega(G) \right\rceil$ colors.
b) Suppose $\omega = 2k$. Then in Procedure 2.2 all together we need at most $3\left\lfloor \frac{\omega(G)}{2} \right\rfloor = \left\lceil \frac{3}{2}\omega(G) \right\rceil$ colors. □

Recall that by definition all vertices of a tetrahedron grid T which are not in G must have weight $d(v) = 0$. Then we need not check whether a vertex of the grid is one of the vertices of G. Therefore:

$$\omega_3(G) = \max\{d(u) + d(v) + d(t) : \{u, v, t\} \in \tau(T)\},$$

where $\tau(T)$ is the set of all triangles of the tetrahedron grid T.

For each vertex $v \in G$, define *base function* κ as

$$\kappa(v) = \max\{a(v, w, t) : \{v, w, t\} \in \tau(T)\},$$

where

$$a(v, w, t) = \left\lceil \frac{d(v) + d(w) + d(t)}{3} \right\rceil,$$

is the rounded average weight of the triangle $\{v, w, t\} \in \tau(T)$.

Clearly, the following fact holds.

Fact 2.3. *For each $v \in G$,*

$$\kappa(v) \leq \left\lceil \frac{\omega_3(G)}{3} \right\rceil \leq \left\lceil \frac{\omega(G)}{3} \right\rceil.$$

We call a vertex v *heavy* if $d(v) > \kappa(v)$, otherwise we call it *light*. If $d(v) > 2\kappa(v)$, we say that the vertex v is *very heavy*.

To color vertices of G we use colors from an appropriate *palette*. For a given color c, its palette is defined as the set of pairs $\{(c, i)\}_{i \in \mathbb{N}}$. A palette is called a *base color palette* if $c \in \{0, 1, 2, 3\}$ is one of the base colors, and it is called an *additional color palette* if $c \notin \{0, 1, 2, 3\}$.

If a vertex v does not have a neighbor of color i in G, we call such color a *free color* of v.

3 Algorithms for multicoloring cannonball graphs

Recall that a tetrahedron grid consists of several horizontal layers which are triangular grids. No matter how we store one layer onto another, for every hexagonal graph in a particular horizontal layer one of the well known algorithms [12, 18, 26] may be used. The best known approximation ratio is $\frac{4}{3}\omega(G')$, where G' is a hexagonal graph in a single layer (obviously $\omega(G') \leq \omega(G)$). Therefore, for each layer we need at most $\frac{4}{3}\omega(G)$ colors. We can use one palette of colors for odd layers and the second palette of colors for even layers, in order to prevent any conflict. Altogether we get an algorithm that uses at most $2 \cdot \frac{4}{3}\omega(G) = \frac{8}{3}\omega(G)$. Since this bound is obviously not the best possible, the algorithm that improves this bound is presented in what follows.

In many papers, e.g. [12, 18, 23, 25], a strategy of borrowing was used. The same idea can be used for cannonball graphs. Our algorithm consists of two main phases. In the first phase (Steps 1 and 2 of the algorithm) vertices take $\kappa(v)$ colors from their base color palette, so use no more than $\frac{4}{3}\omega(G)$ colors. After this phase, all light vertices in G are fully colored, i.e., every light vertex $v \in V(G)$ already received all needed $d(v)$ colors. The vertices that are heavy, but not very heavy, induce a triangle-free cannonball graph with the weighted clique number not exceeding $\lceil \omega(G)/3 \rceil$. Very heavy vertices in G are isolated in the remaining graph and therefore they can easily be fully colored (Step 2 and Step 4). In the second phase (Steps 3 and 4 of the algorithm) we first color all vertices of degree 4 and thereby obtain a 3-colorable graph, for which Procedure 2.2 can be used for satisfying the remaining demands by using new colors.

More precisely, our algorithm consists of the following steps:

Algorithm

Input: A weighted cannonball graph $G = (V, E, d)$. Coordinates $(x(v), y(v), z(v))$, for $v \in V$.

Output: A proper multicoloring of G, using at most $\frac{11}{6} \cdot \omega(G) + \frac{19}{6}$ colors.

Step 0 For each vertex $v \in V$ compute its base color $bc(v)$

and its base function value

$$\kappa(v) = \max\left\{ \left\lceil \frac{d(u) + d(v) + d(t)}{3} \right\rceil : \{v, u, t\} \in \tau(T) \right\},$$

where $\tau(T)$ is the set of all triangles in the tetrahedron grid T.

Step 1 For each vertex $v \in V$ assign $\min\{\kappa(v), d(v)\}$ colors from its base color palette to v. Construct a new weighted triangle-free cannonball graph $G_1 = (V_1, E_1, d_1)$ where $d_1(v) = \max\{d(v) - \kappa(v), 0\}$, $V_1 \subseteq V$ is the set of vertices with $d_1(v) > 0$ (heavy vertices in G) and $E_1 \subseteq E$ is the set of all edges in G with both endpoints in V_1 (G_1 is induced by V_1).

Step 2 For each vertex $v \in V_1$ with $d_1(v) > \kappa(v)$ (very heavy vertices in G) assign the first unused $\kappa(v)$ colors of the base color palettes of its neighbors in the tetrahedron grid T. Construct a new graph $G_2 = (V_2, E_2, d_2)$ where $d_2(v)$ is the difference between $d_1(v)$ and the number of colors assigned in this step. Note that $V_1 = V_2$, and $E_1 = E_2$, as only very heavy vertices are partially colored in this step.

Step 3 For each vertex $v \in V_2$ with $deg_{G_2}(v) = 4$ assign unused colors from its free base color palette. Construct a new 3-colorable graph $G_3 = (V_3, E_3, d_3)$ where $d_3(v)$ is the difference between $d_2(v)$ and the number of colors assigned in this step, $V_3 \subseteq V_2$ is the set of vertices with $d_3(v) > 0$ and $E_3 \subseteq E_2$ is the set of all edges in G_2 with both endpoints in V_3 (G_3 is induced by V_3).

Step 4 Find a 3-coloring of G_3 and apply Procedure 2.2 for the graph G_3 by using colors from new additional color palettes.

4 Correctness proof

Recall that each vertex knows its position on the tetrahedron grid T. Note that whenever we mention “very heavy/heavy/light vertex”, we refer to the property of this vertex in the graph G, i.e., there is no reclassification in graphs $G_i, i \in 1, 2, 3$.

In Step 0 we have to prove that each vertex can obtain its base color. Recall that we can assume that one of the horizontal layers is the base layer. We can compute the base coloring in each vertex $v = (x, y)$ of this layer by the formula: $bc(v) = x \mod 2 + 2(y \mod 2)$. In the neighboring layers (the above and the bottom one) the colors are determined by 4-coloring of the base layer. Thus, we can obtain a proper 4-coloring for the whole cannonball graph. If, in addition we know that G is regular cannonball graph, we can compute the base color directly from expression (2.2).

In Step 1 each heavy vertex v in G is assigned $\kappa(v)$ colors from its base color palette, while each light vertex u is assigned $d(u)$ colors from its base color palette. Hence the remaining weight of each vertex $v \in G_1$ is

$$d_1(v) = d(v) - \kappa(v).$$

Note that G_1 consists only of heavy vertices in G. Therefore,

Lemma 4.1. *G_1 is a triangle-free cannonball graph.*

Proof. Assume that there exists a triangle $\{v, u, t\} \in \tau(G_1)$, which means that $d_1(v)$, $d_1(u), d_1(t) > 0$. Then we have:

$$\begin{aligned} d(v) + d(u) + d(t) &= d_1(v) + \kappa(v) + d_1(u) + \kappa(u) + d_1(t) + \kappa(t) \geq \\ &\geq d_1(v) + d_1(u) + d_1(t) + 3a(u, v, t) \geq \\ &\geq d_1(v) + d_1(u) + d_1(t) + d(v) + d(u) + d(t) \\ &> d(v) + d(u) + d(t) \end{aligned}$$

a contradiction. Therefore, the graph G_1 does not contain a 3-clique, so it is a triangle-free cannonball graph. □

In Step 2 only very heavy vertices in G ($d_1(v) > \kappa(v)$) are colored. It is not difficult to see that each very heavy vertex $v \in G$ is isolated in G_1 (all its neighbors are light in G). Otherwise, for some $\{u, v, t\} \in \tau(T)$, we would have

$$d(v) + d(u) > 2\kappa(v) + \kappa(u) \geq 3a(u, v, t) \geq d(u) + d(v),$$

a contradiction. Let us denote

$$\begin{aligned} D_1(v) &= \min\{\kappa(v) - d(u) : \{u, v\} \in E(T), bc(u) = 1\}, \\ D_2(v) &= \min\{\kappa(v) - d(u) : \{u, v\} \in E(T), bc(u) = 2\}, \\ D_3(v) &= \min\{\kappa(v) - d(u) : \{u, v\} \in E(T), bc(u) = 3\}. \end{aligned}$$

Note that $D_1(v), D_2(v), D_3(v) > 0$. Otherwise, we would have let say $D_1(v) \leq 0$ and thus $d(u) \geq \kappa(v)$ for some u such that $\{u, v, t\} \in \tau(T)$. Therefore,

$$d(v) + d(u) > 2\kappa(v) + \kappa(v) = 3\kappa(v) \geq 3a(u, v, t) \geq d(u) + d(v),$$

a contradiction.

Since in Step 1 each light vertex t uses exactly $d(t)$ colors from its base color palette, very heavy vertex v have at least $D_i(v)$ free colors from every base color palette i. Besides, for heavy neighbors $\{u, v\} \in E$, we can prove:

Lemma 4.2. *In G_1 for every edge $\{v, u\} \in E_1$ we have:*

$$d_1(v) + d_1(u) \leq \kappa(v), \quad d_1(u) + d_1(v) \leq \kappa(u).$$

Proof. Assume that v and u are heavy vertices in G and $d_1(v) + d_1(u) > \kappa(v)$. Then for some $\{v, u, t\} \in \tau(T)$ we have:

$$d(v)+d(u) = d_1(v)+\kappa(v)+d_1(u)+\kappa(u) > 2\kappa(v)+\kappa(u) \geq 3a(u, v, t) \geq d(u)+d(v),$$

a contradiction. □

Another useful observation is

Claim 4.3.

$$\omega(G_2) \leq \left\lceil \frac{\omega(G)}{3} \right\rceil .$$

Proof. Recall that in a cannonball graph the only cliques are tetrahedrons, triangles, edges and isolated vertices. Since G_1 is a triangle-free cannonball graph, G_2 contains no tetrahedron, neither triangle, so we have only edges and isolated vertices to check.

For each edge $vu \in E_2$, using Lemma 4.2 and Fact 2.3, we have:

$$d_2(v) + d_2(u) \leq d_1(v) + d_1(u) \leq \kappa(v) \leq \lceil \omega(G)/3 \rceil.$$

For each isolated vertex $v \in G_2$ we should have $d_2(v) \leq \lceil \omega(G)/3 \rceil$. If v is very heavy, then $d_2(v) = d(v) - 2\kappa(v)$ because the vertex has received $\kappa(v)$ colors in Step 1 and in Step 2. We claim that $d_2(v) \leq \kappa(v)$. Indeed, if $d_2(v) > \kappa(v)$, then $d(v) = d_2(v) + 2\kappa(v) > 3\kappa(v)$ contradicting the definition of $\kappa(v)$. Hence, $d_2(v) \leq \kappa(v) \leq \lceil \omega(G)/3 \rceil$ as needed. If v is not very heavy in G then $\kappa(v) < d(v) \leq 2\kappa(v)$ and $d_2(v) = d(v) - \kappa(v) \leq \kappa(v) \leq \lceil \omega(G)/3 \rceil$. □

Let $\Delta(G)$ be the maximal vertex degree in the graph G. Considering correctness of Step 3, we have to first prove that:

Lemma 4.4. *$\Delta(G_2) \leq 4$ and every vertex v with $deg_{G_2}(v) = 4$ has at least one free color.*

Proof. Let v be an arbitrary vertex in the graph G_2. Without loss of generality, assume that $bc(v) = 0$. Recall that by Lemma 4.1 graph G_2 is triangle-free. Therefore, vertex v can have at most 3 neighbors in its layer, and the angle between any two of them is $2\pi/3$. In this case vertex v cannot have any additional neighbor in the lower or in the upper layer, therefore $deg_{G_2}(v) = 3$. If the vertex v has only 2 neighbors in its layer, then we have two different possibilities for the angle between the neighbors: $2\pi/3$ and π. Both possibilities on the layer-arrangement (a) are depicted on Figure 3 and the (b) case is depicted in Figure 4. It is easy to see that vertex v could have at most two additional neighbors in lower and upper layer - otherwise we obtain a triangle. Suppose that $deg_{G_2}(v) = 4$, then all possible cases of its neighbourhood are shown in Figures 3 and 4. It is easy to see that in both cases (1) vertex v can borrow color 1, and in both cases (2) vertex v can borrow color 2 or 3. If the vertex v has only one neighbor in its layer it can have at most three neighbors altogether. Namely, in this case v can have at most one neighbor in the lower layer and at most one neighbor in the upper layer, since otherwise we obtain a triangle. □

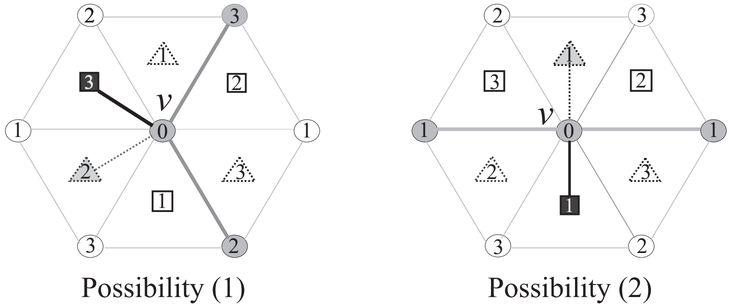


Figure 3: Two different possibilities for neighbourhood of vertex v with $deg_{G_2}(v) = 4$ in a triangle-free cannonball graph, obtained from the layer-arrangement (a). Circles represent vertices of the middle layer, squares of the upper layer and triangles of the lower layer, and white vertices are part of the grid, but are not in the graph.

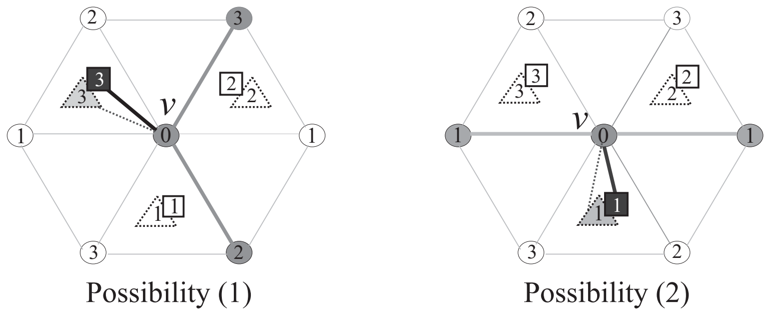


Figure 4: Two different possibilities for vertex v neighbourhood with $deg_{G_2}(v) = 4$ in a triangle-free cannonball graph, obtained from the layer-arrangement (b). Notation has the same meaning as in Figure 3.

By Lemma 4.4 we know that borrowing is possible for all vertices of degree 4.

In Step 3 we take the colors from the free base color palettes.

Without loss of generality, assume that $bc(v) = 0$ and one of its free colors is 1. Recall that in this case we have $D_1(v)$ free colors from the first base color palette, which is enough to fully color the vertex v (current demand $d_2(v)$). Namely,

Lemma 4.5.

$$d_2(v) \leq D_1(v).$$

Proof. Let v be a vertex in G_2 with $bc(v) = 0$. If vertex $v \in G_2$ has four neighbors in G_2, it always has free colors such that all neighbors on its layer of this base colors are not in $V(G_2)$ (see Figures 3 and 4). Without loss of generality assume that one of these free colors is 1. Let t be a vertex, which is an existing neighbor of v in G_2, and u is the neighbor of v with $bc(u) = 1$ so that $\{u, v, t\} \in \tau(T)$ is a triangle. Then we have

$$\kappa(v) + d_2(v) + a(u, v, t) + d(u) \overset{\star}{\leq} d(v) + d(t) + d(u) \leq 3a(u, v, t) \leq a(u, v, t) + 2\kappa(v)$$

and the inequality $\star$ occurs because $d_2(v) = d(v) - \kappa(v)$ and $d(t) > \kappa(t) \geq a(u, v, t)$. Therefore,

$$d_2(v) \leq \kappa(v) - d(u) \leq D_1(v).$$

□

Finally, to show correctness of Step 4 we have to prove that for G_3 we can apply Procedure 2.2. We know that $\Delta(G_3) \leq 3$ since $\Delta(G_2) \leq 4$ and in Step 3 we had fully colored all vertices with degree equal to 4. According to Brooks' Theorem [4] we know that G_3 is 3-colorable. It is well-known that the 3-coloring can be found in polynomial time [11]. In fact a linear time algorithm exists [3].

Therefore, we can apply Procedure 2.2 and multicolor G_3 by using $\lceil \frac{3}{2}\omega(G_3) \rceil - 1$ new colors.

Ratio and time complexity

We claim that during the first three steps our algorithm uses at most $\frac{4}{3}\omega(G) + \frac{8}{3}$ colors. To see this, notice that in Step 1 each vertex v uses at most $\kappa(v)$ colors from its base color palette and, by Fact 2.3 and using that there are four base colors, we know that no more than $4\lceil \omega(G)/3 \rceil \leq \frac{4}{3}\omega(G) + \frac{8}{3}$ colors are needed. Note also that in Step 2 and Step 3 we use only those colors from the base color palettes which were not used in Step 1, so altogether no more than $\frac{4}{3}\omega(G) + \frac{8}{3}$ colors from the base color palettes are used in total until Step 4.
In Step 4 we introduce new palettes that contain no more than $\lceil \frac{3}{2}\omega(G_3) \rceil$ colors (by Lemma 2.1) .

Let $A(G)$ denote the number of colors used by our algorithm for the graph G. By Claim 4.3 it holds $\omega(G_3) \leq \omega(G_2) \leq \lceil \omega(G)/3 \rceil \leq \omega(G)/3 + \frac{2}{3}$ and thus

$$\left\lceil \frac{3}{2}\omega(G_3) \right\rceil \leq \left\lceil \frac{3}{2}\left(\frac{\omega(G)}{3} + \frac{2}{3}\right) \right\rceil = \left\lceil \frac{\omega(G)}{2} + 1 \right\rceil \leq \frac{1}{2}\omega(G) + \frac{3}{2}.$$

Therefore, the total number of colors used by our algorithm is at most

$$A(G) \leq \frac{4}{3}\omega(G) + \frac{8}{3} + \left\lceil \frac{3}{2}\omega(G_3) \right\rceil \leq \frac{4}{3}\omega(G) + \frac{8}{3} + \frac{1}{2}\omega(G) + \frac{3}{2} = \frac{11}{6}\omega(G) + \frac{25}{6}.$$

Hence we arrived at the statement of Theorem 1.1.

Finally, we wish to remark that our algorithm can be implemented in linear time. Namely, in Steps 1, 2 and 3 we need constant time for each vertex. In Step 0 we need linear time to compute the value of κ. Step 4 is also linear since we have linear 3-coloring algorithm [3] and Procedure 2.2 is constant.

5 Conclusion

In this paper we provide an algorithm for a proper multicoloring of a cannonball graph that uses at most $\frac{11}{6}\omega(G) + \frac{19}{6}$ colors. As this is the first result for the multicoloring problem of cannonball graphs, we believe that further improvements can be done. Among the interesting problems that remain open are: improving of the competitive ratio $11/6$, finding some distributed algorithms for multicoloring cannonball graphs, or finding some k-local algorithms for some k, similarly as in 2-dimensional case for hexagonal graphs (for definition of k-local algorithms see [10]). We already mentioned that in the 2-dimensional case, better bounds were obtained for triangle-free hexagonal graphs. It is very likely that also for cannonball graphs there exist some "forbidden" subgraphs H, maybe tetrahedrons, such that better bounds can be obtained for H-free cannonball graphs.

Acknowledgement. The authors wish to thank to the anonymous referee for very careful reading and for constructive remarks which helped to considerably improve the presentation.

References

[1] K. Aardal, S van Hoesel, A. Koster, C. Mannino and A.Sassano, Models and solution techniques for frequency assignment problems, *Ann. Oper. Res.* **153** (2007), 79–129.

[2] L. G. Anderson, A simulation study of some dynamic channel assignment algorithms in a high capacity mobile telecommunication system, *IEEE Transaction on Communications* **21** (1973), 1294–1301.

[3] B. Baetz and D. R. Wood: *Brooks' Vertex-Colouring Theorem in Linear Time*, arXiv: 1401.8023v1[cs.DM] 30 Jan 2014 B.

[4] R. L. Brooks, On coloring the nodes of a network, *Proceedings of the Cambridge Philosophical Society* **37 (2)** (1941), 118–121.

[5] W.K. Hale, Frequency assignment: theory and applications, *Proceedings of the IEEE* **68 (12)** (1980), 1497–1514.

[6] T. Hales, Cannonballs and honeycombs, *Notices of the American Mathematical Society* **47** (2000), 440–449.

[7] T. Hales, A proof of the Kepler conjecture, *Annals of Mathematics. Second Series* **162** (2005), 1065–1185.

[8] F. Havet, Channel assignment and multicoloring of the induced subgraphs of the triangular lattice, *Discrete Mathematics* **233** (2001), 219–231.

[9] F. Havet, J. Žerovnik, Finding a Five Bicolouring of a Triangle-free Subgraph of the Triangular Lattice, *Discrete Mathematics* **244** (2002),103–108.

[10] J. Janssen, D. Krizanc, L. Narayanan, S. Shende, Distributed Online Frequency Assignment in Cellular Network, *Journal of Algorithms* **36 (2)** (2000), 119–151.

[11] L. Lovasz, Three short proofs in graph theory, *Journal of Combinatorial Theory, Series B* **19** (1975), 269–271.

[12] C. McDiarmid, B. Reed, Channel assignment and weighted coloring, *Networks* **36 (2)** (2000), 114–117.

[13] L. Narayanan, *Channel assignment and graph multicoloring, Handbook of wireless networks and mobile computing*, Wiley, New York, 2002, 71–94.

[14] L. Narayanan, S.M. Shende, Static frequency assignment in cellular networks, *Algorithmica* **29 (3)** (2001), 396–409.

[15] I. Sau, P. Šparl, J. Žerovnik, 7/6-approximation Algorithm for Multicoloring Triangle-free Hexagonal Graphs, *Discrete Mathematics* **312** (2012), 181–187.

[16] P. Šparl, R. Witkowski, J. Žerovnik, A Linear Time Algorithm for $7-[3]$-coloring Triangle-free Hexagonal Graphs, *Information Processing Letters* **112** (2012), 567–571.

[17] P. Šparl, R. Witkowski, J. Žerovnik, 1-local 7/5-Competitive Algorithm for Multicoloring Hexagonal Graphs, *Algorithmica* **64** (2012), 564–583.

[18] P. Šparl, J. Žerovnik, 2-local 4/3-competitive Algorithm for Multicoloring Hexagonal Graphs, *Journal of Algorithms* **55 (1)** (2005), 29–41.

[19] P. Šparl, J. Žerovnik, 2-local 5/4-competitive algorithm for multicoloring triangle-free hexagonal graphs, *Information Processing Letters* **90 (5)** (2004), 239–246.

[20] P. Šparl, J. Žerovnik, 2-local 7/6-competitive algorithm for multicoloring a sub-class of hexagonal graph, *International Journal of Computer Mathematics* **87** (2010), 2003–2013.

[21] K.S. Sudeep, S. Vishwanathan, A technique for multicoloring triangle-free hexagonal graphs, *Discrete Mathematics* **300** (2005), 256–259.

[22] B. H. Yandell, *The Honors Class. Hilbert's Problems and Their Solvers*, A K Peters, 2002.

[23] R. Witkowski, A 1-local 17/12-competitive Algorithm for Multicoloring Hexagonal Graphs, *Lecture Notes of Computer Science* **5699/2009** (2009), 346–356.

[24] R. Witkowski, 1-local 4/3-competitive Algorithm for Multicoloring a Subclass of Hexagonal Graphs, in press in *Discrete Applied Mathematics* (2012), DOI: 10.1016/j.dam.2011.12.013

[25] R. Witkowski, J. Žerovnik, 1-local 7/5-competitive Algorithm for Multicoloring Hexagonal Graphs, *Electronic Notes in Discrete Mathematics* **36** (2010), 375–382.

[26] R. Witkowski, J. Žerovnik, 1-local 33/24-competitive Algorithm for Multicoloring Hexagonal Graphs, *Discrete Mathematics & Theoretical Computer Science* **15 (3)** (2013), 127–138.

[27] J. Žerovnik, A distributed 6/5-competitive algorithm for multicoloring triangle-free hexagonal graphs, *International Journal of Pure and Applied Mathematics* **23 (2)** (2005), 141–156.

[28] Algorithmische Graphentheorie, Oberwolfach, December 8-14, 2002. (Report No. 55/2002). Oberwolfach: Mathematisches Forschungsinstitut, 2002. (seminar page `http://www.mfo.de/occasion/0250/www\_view`, link to report `www.mfo.de/document/0250/Report55\_2002.ps`)

[29] FAP website, `http://fap.zib.de/`

ARS MATHEMATICA CONTEMPORANEA

Also available at http://amc-journal.eu
ISSN 1855-3966 (printed edn.), ISSN 1855-3974 (electronic edn.)
ARS MATHEMATICA CONTEMPORANEA 10 (2016) 45–66

Distinguishing partitions of complete multipartite graphs

Michael Goff

Vanderbilt University, Department of Mathematics,
1326 Stevenson Center, Nashville, USA

Received 2 January 2013, accepted 23 February 2014, published online 27 May 2015

Abstract

A *distinguishing partition* of a set X with automorphism group $\mathrm{aut}(X)$ is a partition of X that is fixed by no nontrivial element of $\mathrm{aut}(X)$. In the event that X is a complete multipartite graph with its automorphism group, the existence of a distinguishing partition is equivalent to the existence of an asymmetric hypergraph with prescribed edge sizes. An asymptotic result is proven on the existence of a distinguishing partition when X is a complete multipartite graph with m_1 parts of size n_1 and m_2 parts of size n_2 for small n_1, m_2 and large m_1, n_2. A key tool in making the estimate is counting the number of trees of particular classes.

Keywords: Complete multipartite graph, distinguishing partition, combinatorial species, tree enumeration.

Math. Subj. Class.: 05C25, 05C65, 20B25

1 Introduction

The distinguishing partition problem asks, given a finite set X with a group $\mathcal{G}$ that acts on X, whether there exists a partition P of the elements of X such that no nontrivial element of $\mathcal{G}$ fixes P. Formally, consider a partition $P = \{P_1, \ldots, P_t\}$ and $\gamma \in \mathcal{G}$. For general $X' = \{x_1, \ldots, x_i\} \subset X$, let $\gamma(X') = \{\gamma(x_1), \ldots, \gamma(x_i)\}$. Then let $\gamma(P) = \{\gamma(P_1), \ldots, \gamma(P_t)\}$. We say that P is a distinguishing partition if $\gamma(P) \neq P$ for all nontrivial $\gamma \in \mathcal{G}$. When X is a graph, we consider it to be acted upon by its automorphism group $\mathrm{aut}(X)$.

Not all sets X with group action $\mathcal{G}$ have a distinguishing partition. For example, if $\mathcal{G}$ is the group of all permutation on X and $|X| \geq 2$, then X does not have a distinguishing partition. Conversely, if $\mathcal{G}$ is the trivial group, then all partitions of X are distinguishing. As another example, let X be the set $\{a, b, c, d\}$ acted upon by the cyclic group with

E-mail address: michael.goff@vanderbilt.edu (Michael Goff)

generator that takes a to b, b to c, c to d, and d to a. Then X has the following distinguishing partitions: $\{\{a\},\{b,c,d\}\}, \{\{b\},\{a,c,d\}\}, \{\{c\},\{a,b,d\}\}, \{\{d\},\{a,b,c\}\}, \{\{a,b\},\{c\},\{d\}\}, \{\{b,c\},\{d\},\{a\}\}, \{\{c,d\},\{a\},\{b\}\}, \{\{d,a\},\{b\},\{c\}\}$. By contrast, the dihedral group acting on four elements has no distinguishing partition.

In general, the conditions for the existence of a distinguishing partition can be quite complex, even in a relatively restricted setting such as taking X to be a complete multipartite graph, acted upon by its automorphism group. Informally, the difficulty is that if a partition P consists of few large parts, then a nontrivial automorphism might fix each part, while if P consists of many small parts, then a nontrivial automorphism might permute the parts.

Ellingham and Schroeder [7] first considered the distinguishing partitions problem for complete equipartite graphs. Their finding is that if X is a complete equipartite graph with m parts, each of size n, then X has a distinguishing partition if and only if $m \geq f(n)$ for $f(2) = f(14) = 6$, $f(6) = 5$, and otherwise $f(n) = \lfloor \log_2(n+1) \rfloor + 2$. In this setting, $\mathrm{aut}(X)$ is the imprimitive action of the wreath product $S_n \wr S_m$ on X.

The distinguishing partition is a measure of the level of symmetry of a group action, and as such the concept is closely related to the well-studied distinguishing number, as introduced by Albertson and Collins [1] on a graph and by Tymoczko [11] for a general group action. Other such measures are the cost of 2-distinguishing [6] and the determining set [5]. The survey of Bailey and Cameron [2] shows how these concepts have appeared independently in many different settings.

The distinguishing number of X is the mimimum number of label classes in a *distinguishing labeling* of X. In turn, a distinguishing labeling is a map from X to the set of labels $[t]$ that is not fixed under any nontrivial automorphism of X. All distinguishing partitions can be regarded as distinguishing labelings by treating each block of the partition as a separate label class, but not all distinguishing labelings are similarly distinguishing partitions. Every set with group action has a distinguishing labeling–every element could be assigned a unique label–but not all have a distinguishing partition. It should be noted that the term "distinguishing partition" has been used elsewhere to mean what we here call a distinguishing labeling.

For the remainder of this paper, we will consider the case that X is a complete multipartite graph with its automorphism group. We denote by $X = K_{n_1,\ldots,n_m}$ the complete multipartite graph with maximal independent sets X_i of size n_i for $1 \leq i \leq m$. Also, $K_{m_1(n_1),m_2(n_2)}$ denotes the complete multipartite graph with m_i parts each of size n_i for $i = 1, 2$. We focus in particular on $K_{m_1(n_1),m_2(n_2)}$ for fixed n_1 and m_2 and large m_1 and n_2.

Based on the results of Ellingham and Schroeder [7], we might expect a complete multipartite graph to have a distinguishing partition if it has many small parts, and not to have a distinguishing partition if it has few large parts. In our setting, which combines these two extremes, it seems natural to expect that a distinguishing partition, in the asymptotic sense, would exist if n_2/m_1 does not exceed a certain ratio. Our main result is that this is indeed the case.

Theorem 1.1. *Fix $n_1 \geq 2$ and $m_2 \geq 1$, and suppose that m_1 is sufficiently large relative to n_1 and m_2. There exists a value $r = r_{n_1,m_2}$ such that the following holds. $K_{m_1(n_1),m_2(n_2)}$ has a distinguishing partition if and only if*

$$n_2 \leq r_{n_1,m_2} m_1 + \epsilon(m_1)$$

for some function $\epsilon(m_1) \in o(m_1)$.

We have that $r_{2,m_2} = 1$. For $n_1 \geq 3$, we define r_{n_1,m_2} by first choosing values of $j = j_{n_1,m_2}$ and $k = k_{n_1,m_2}$ such that

$$n_1 = 2 + \binom{m_2}{0} + \binom{m_2}{1} + \cdots + \binom{m_2}{j} + k,$$

with either

$$j < \lfloor (m_2-1)/2 \rfloor \quad \text{and} \quad 0 \leq k < \binom{m_2}{j+1}, \quad \text{or} \quad j = \lfloor (m_2-1)/2 \rfloor \quad \text{and} \quad k \geq 0.$$

If $j < \lfloor (m_2-1)/2 \rfloor$, then let

$$r = 1 + \sum_{i=0}^{j} \frac{m_2 - i}{m_2}\binom{m_2}{i} + \frac{m_2 - j - 1}{m_2}k,$$

and otherwise choose

$$r = 1 + \sum_{i=0}^{j} \frac{m_2 - i}{m_2}\binom{m_2}{i} + \frac{1}{2}k.$$

We say that $j_{2,m_2} = -1$.

The structure of the paper and the proof Theorem 1.1 is as follows. In Section 2, we establish basic concepts on enriched trees and hypergraphs which are used heavily throughout the proof. In Section 3, we show how a type of partition of $K_{m_1(n_1),m_2(n_2)}$ known as a regular partition may be represented as a hypergraph with m_i edges of size n_i, $i = 1, 2$. We establish key lemmas for the general result in Section 4. In Section 5, we provide the general construction that, for the existence of a distinguishing partition, maximizes n_2 to within an additive constant, given m_1, n_1, m_2. Then we prove that for large m_1 relative to n_1 and m_2, if $n_2' > n_2$ and $K_{m_1(n_1),m_2(n_2')}$ has a distinguishing partition, then so does $K_{m_1(n_1),m_2(n_2)}$.

In Section 6, we focus on the case that $n_1 = 2$. Then the following refinement of Theorem 1.1 holds.

Theorem 1.2. *There exist constants* $\alpha > 0$ *and* $\beta > 1$ *and*

$$z := \left\lfloor \log_\beta \left(\frac{m_1(\beta-1)}{\alpha\beta} \left(\log_\beta m_1\right)^{3/2} \right) \right\rfloor$$

such that Theorem 1.1 holds with $\epsilon(m_1)$ *of the form*

$$\frac{m_1}{z+1} + (1 + o_{m_1}(1))\alpha\beta^z z^{-7/2}\left(\frac{\beta}{\beta-1}\right)^2 \approx \frac{m_1}{\log_\beta(m_1)}.$$

In Section 7, we consider the case that $k = 0$ and $j < \lfloor (m_2-1)/2 \rfloor$. Then Theorem 1.1 can be refined as follows.

Theorem 1.3. *If* $k = 0$ *and* $j < \lfloor (m_2-1)/2 \rfloor$*, then Theorem 1.1 holds with* $\epsilon(m_1)$ *of the form*

$$\left(\frac{(2m_2 - 4j - 4)^{\frac{2m_2-4j-5}{2m_2-4j-4}}}{2m_2 - 4j - 5} C^{\frac{1}{2m_2-4j-4}} + o_{m_1}(1) \right) m_1^{\frac{2m_2-4j-5}{2m_2-4j-4}}$$

for a value of C *that depends only on* n_1 *and* m_2.

The value of C will be specified in Section 7.

We consider $k \geq 1$ and $j < \lfloor (m_2 - 1)/2 \rfloor$ in Section 8.

Theorem 1.4. *If $k \geq 1$ and $j < \lfloor (m_2 - 1)/2 \rfloor$, then Theorem 1.1 holds with $\epsilon(m_1)$ of the form $\Theta(m_1/(\log m_1))$.*

In Section 9, we consider the case that $k = 0$ and $j = (m_2 - 2)/2$. Then the following exact result for large m_1 is possible.

Theorem 1.5. *Suppose that $k = 0$ and $j = \lfloor (m_2 - 1)/2 \rfloor$. Theorem 1.1 holds with $\epsilon(m_1) = 2^{m_2 - 1}$ if m_2 is even and at least 4 and $\epsilon(m_1) = 2^{m_2-1} - 1$ if m_2 is odd or 2, for sufficiently large m_1.*

In Section 10, we prove the following for $k \geq 1$ and $j = \lfloor (m_2 - 1)/2 \rfloor$.

Theorem 1.6. *If $k \geq 1$ and $j = \lfloor (m_2 - 1)/2 \rfloor$, then Theorem 1.1 holds with $\epsilon(m_1) = 2^{m_2-1} - 1$ if km_1 is even and m_2 is odd, and otherwise $\epsilon(m_1) = 2^{m_2-1} + \lfloor rm_1 \rfloor - rm_1$, for sufficiently large m_1.*

2 Enriched trees and hypergraphs

Combinatorial species and enriched trees

We make use of the language of combinatorial species, as presented by Bergeron, Labelle, and Leroux [4]. A *species* F is, for every finite set U, a finite set of objects $F[U]$, called structures, together with, for every bijection $\sigma : U \to U'$, a function $F[\sigma] : F[U] \to F[U']$ that satisfies the following two properties, which are standard functioriality properties in category theory:

1) for all bijections $\sigma : U \to U'$ and $\sigma' : U' \to U''$, $F[\sigma' \circ \sigma] = F[\sigma'] \circ F[\sigma]$,
2) for the identity map Id_U, $F[\mathrm{Id}_U] = \mathrm{Id}_{F[U]}$.

The function $F[\sigma]$ is known as transport of species. Consider the symmetric group S_U that acts on U. Given an F-structure s, we say that the automorphism group of s, $\mathrm{aut}(s)$, is the subgroup of those $\sigma \in S_U$ that satisfy $F[\sigma](s) = s$.

Let $\mathfrak{a}$ be the species of asymmetric trees, or trees whose automorphism group is trivial, and let $\mathfrak{a}^\bullet$ be the species of rooted asymmetric trees. A rooted tree is considered asymmetric if it has no nontrivial root-preserving automorphism; it is possible that the underlying unrooted tree structure is not asymmetric.

Now let F be a species that contains at least one structure over a set of size 1. An F-*enriched tree* on a set U is a tree on U together with an F-structure s_v on the neighbor set $N(v)$ of every vertex $v \in U$. If σ is an automorphism of an F-enriched tree t, then σ is an automorphism of the underlying tree structure of t. Furthermore, if σ_v is the restriction of σ on $N(v)$, then the transport of species $F[\sigma_v]$ takes s_v to $s_{\sigma(v)}$. We say that $\mathfrak{a}_F$ is the species of asymmetric F-enriched trees. For example, when F is $\mathcal{E}$, the species of sets, then there is a unique $\mathcal{E}$-structure on every finite set, and $\mathfrak{a}_\mathcal{E}$ is simply $\mathfrak{a}$. The species $\mathfrak{A}$ and $\mathfrak{A}_F$ are, respectively, the species of (not necessarily asymmetric) trees and the species of F-enriched trees.

The sum of two species $(F + F')[U]$ is the disjoint union $F[U] + F'[U]$ such that $(F + F')[\sigma](s) = F[\sigma](s)$ if $s \in F[U]$ and $(F + F')[\sigma](s) = F'[\sigma](s)$ if $s \in F'[U]$. If a

is a positive integer, then aF is the sum of a copies of F. We say that $F_i[U]$ is $F[U]$ when $|U| = i$, and otherwise $F_i[U] = \emptyset$, and $|F_i|$ is the number of structures of F over a set with i elements.

Consider the species $F = \sum_{i=1}^{\kappa} a_i \mathcal{E}_i$ for nonnegative integers $\kappa, a_1, \ldots, a_\kappa$ with $\kappa \geq 1$ and $a_1 \geq 1$. This is the species that consists of a_i distinct set structures over a set with i elements for $1 \leq i \leq \kappa$ and otherwise no structures. Then the species $\mathfrak{a}_F$ may be regarded as the species of asymmetric trees in which every vertex has degree at most κ, and every vertex of degree i is assigned a label from a pool of a_i possible labels. Such a tree is asymmetric if it has no nontrivial label-preserving automorphism. Later, we show how estimating the number of elements of $\mathfrak{a}_F$ with a given number of vertices can help determine asymptotic bounds for the distinguishing partitions problem.

A structure of the product species FF' over U is an ordered pair (f, f') for an F-structure f over U_1 and a F'-structure f' over U_2 for some partition $U_1 \sqcup U_2$ of U. Transport of species is defined by $(FF')[\sigma](s) = (F[\sigma_1](f), F'[\sigma_2](f'))$, each σ_i the restriction of σ to U_i.

Hypergraphs

A *hypergraph* H is a triple $(V(H), E(H), I(H))$, with $V(H)$ a finite set of elements called *vertices* and $E(H)$ a finite set of elements called *edges*. The *incidence relation* $I(H)$ is a subset of $V(H) \times E(H)$. We will generally treat edges as subsets of $V(H)$. The *degree* of $v \in V(H)$, denoted $\deg(v)$, is the number of edges incident to v. H is connected if the bipartite graph with vertex sets $V(H)$ and $E(H)$ and edge set $I(H)$ is connected, and $E(H) \neq \emptyset$. If every edge is incident to n_1-vertices, then H is n_1*-uniform.*

For a hypergraph H with an edge e, $\deg_1(H)$ and $\deg_1(e)$ denote the number of vertices of degree 1 in H and e, while $\deg_2^+ H$ and $\deg_2^+ e$ are the numbers of vertices of degree at least 2 in H and e.

An automorphism σ of H is a permutation of $V(H)$ and $E(H)$ such that $(v, e) \in I(H)$ if and only if $(\sigma(v), \sigma(e)) \in I(H)$ for all vertices v and edges e. We say that σ is *trivial* if $\sigma(v) = v$ and $\sigma(e) = e$ for all vertices v and edges e, and H is *asymmetric* if the only automorphism of H is trivial. Thus we allow that hypergraphs may contain multiple edges that are incident to the same vertex set, but such hypergraphs are not asymmetric.

A connected n_1-uniform hypergraph H is called a *tree* if $E(H)$ can be enumerated $\{e_1, \ldots, e_{|E(H)|}\}$ in such a way that for each $2 \leq i \leq |E(H)|$, we have that $|e_i \cap (e_1 \cup \cdots \cup e_{i-1})| = 1$. Equivalently, a tree is a connected n_1-uniform hypergraph H with $(n_1 - 1)|E(H)| + 1$ vertices. The *leaves* of a tree H are the edges e that satisfy $\deg_1(e) = n_1 - 1$. We say that $l(G)$ is the number of leaves of G.

For a tree G, define the quantity

$$\mu(G) := \sum_{\substack{v \in V(G) \\ \deg(v) > 2}} (\deg(v) - 2).$$

We will later need the following relationship between $l(G)$ and $\mu(G)$.

Lemma 2.1. *Let G be a tree with at least* 2 *edges. Then* $l(G) \geq \mu(G) + 2$.

Proof. Enumerate $E(G) = (e_1, \ldots, e_{|E(G)|})$ such that for $2 \leq i \leq |E(G)|$,

$$e_i \cap (e_1 \cup \ldots \cup e_{i-1}) = \{v_i\},$$

and let G_i be the subtree of G with edges $(e_1, \dots, e_i)$. We prove that the lemma holds for G_i by induction on i for $2 \le i \le |E(G)|$, with the $i = 2$ case following from $\mu(G_2) = 0$ and $l(G_2) = 2$. Assume the lemma holds for G_{i-1}.

For $3 \le i \le |E(G)|$, $\mu(G_i) = \mu(G_{i-1}) + 1$ if v_i has degree at least 3 in G_i, and otherwise $\mu(G_i) = \mu(G_{i-1})$. Also, $l(G_i) \ge l(G_{i-1})$ since e_i is a leaf in G_i, and at most one leaf of G_{i-i}, namely a leaf that contains v_i as a degree 1 vertex, is not a leaf in G_i. Furthermore, whenever $\mu(H_i) = \mu(G_{i-1}) + 1$, v_i has degree at least 2 in G_{i-1} and thus $l(G_i) = l(G_{i-1}) + 1$. The lemma follows for G_i by $l(G_i) - l(G_{i-1}) \ge \mu(G_i) - \mu(G_{i-1})$. □

3 Distinguishing partitions and asymmetric hypergraphs

We demonstrate a bijection between certain distinguishing partitions of complete multipartite graphs and asymmetric hypergraphs with prescribed edge sizes. Using this bijection, we establish the existence or nonexistence of distinguishing partitions by demonstrating the existence or nonexistence of certain asymmetric hypergraphs. The argument is nearly identical to that given by Ellingham and Schroeder [7].

With $X = K_{n_1,\dots,n_m}$, let P be a partition of X with parts $P_1, \dots, P_t$, and we say that P is a *regular partition* of X if $|X_i \cap P_{i'}| \le 1$ for all i and i'. It is a necessary but not sufficient condition for P to be distinguishing that P is regular.

Definition 3.1. For every regular partition P of X with parts $P_1, \dots, P_t$, we associate a hypergraph $\tau(P)$ as follows: $V(\tau(P)) = \{P_i, 1 \le i \le t\}$, $E(\tau(P)) = \{X_i : 1 \le i \le m\}$, and X_i and $P_{i'}$ are incident if $|X_i \cap P_{i'}| = 1$.

Note that $\tau(P)$ is a hypergraph with m edges with sizes $n_1, \dots, n_m$ since X_i intersects exactly n_i parts of P.

We say that the automorphism group aut(P) is the subgroup of aut(X) consisting of those elements that fix P. The following relationship holds.

Lemma 3.2. *If P is a regular partition of X, then* aut(P) *is isomorphic to* aut$(\tau(P))$.

Proof. Let $\tilde{\tau}$: aut$(P) \to$ aut$(\tau(P)$ be the group homomorphism induced by τ. Say that $P_1, \dots, P_t$ are the parts of P.

An automorphism $\sigma \in$ aut(P) induces automorphisms σ_P and σ_X on the sets $\{P_i\}$ and $\{X_{i'}\}$ respectively. Then σ is uniquely determined by σ_P and σ_X, and in particular σ is trivial if and only if σ_P and σ_X are both trivial. Thus $\tilde{\tau}$ is injective.

Now let $\sigma' \in$ aut$(\tau(P))$. Then σ' is uniquely determined by incidence-preserving permutations of $\{P_i\}$ and $\{X_{i'}\}$. Let σ be the permutation of X such that if $x \in X$ is the unique vertex contained in $X_i \cap P_{i'}$, then $\sigma(x)$ is the unique vertex contained in $\sigma'(X_i) \cap \sigma'(P_{i'})$. It is readily checked that in fact $\sigma \in$ aut(P), and thus $\tilde{\tau}$ is surjective. □

Corollary 3.3. *There exists a distinguishing partition of $K_{n_1,\dots,n_m}$ if and only if there exists an asymmetric hypergraph with m edges of sizes $n_1, \dots, n_m$.*

It will be convenient to associate another hypergraph with a regular partition P of $K_{m_1(n_1),m_2(n_2)}$. Let $\tau'(P)$ be a vertex-labeled hypergraph that contains exactly the vertices and the n_1-edges of $\tau(P)$. Say that the n_2-edges of $\tau(P)$ are $X_1, \dots, X_{m_2}$. Then the vertex label set of $\tau'(P)$ is $2^{[m_2]}$, and a vertex v in $\tau'(P)$ is labeled with a set $S \subseteq [m_2]$ if $v \in X_i$ exactly when $i \in S$. Then $\tau'(P)$ is just a different way of encoding $\tau(P)$.

The *weight* $w(v)$ of a vertex $v \in \tau'(P)$ is the cardinality of its label. The *weight* $w(S)$ of a set of vertices S is the sum of the weights of the vertices in S. The *weight* $w(e)$ or $w(G)$ of an n_1-edge e or connected component $G \subset \tau'(P)$ is the weight of the vertex set of e or G. The *value* of G is $w(G) - rm_2|E(G)|$. Value may be positive or negative. We have that $n_2 = w(\tau'(P))/m_2$, and thus our strategy in proving the main results is to find an asymmetric labeled hypergraph with m_1 n_1-edges and maximal weight. Though weight and value encode the same information, value is useful in that it gives a clear comparison of the weight of a component of $\tau'(P)$ to its asymptotic limit.

4 Key Lemmas

We now present a series of lemmas that provide upper bounds on the weights and values of certain types of components.

Lemma 4.1. *Let G be a connected n_1-uniform hypergraph with $n_1|E(G)|/2 + p$ vertices. Then G contains $2p + \mu(G)$ vertices of degree 1. Equivalently, each edge has, on average, $(2p + \mu(G))/|E(G)|$ degree 1 vertices.*

Proof. There are $n_1|E(G)|$ pairs of the form (v, e), where v is a vertex, e an edge, and $v \in e$. The number of such pairs (v, e) is also

$$\deg_1(G) + 2(n_1|E(G)|/2 + p - \deg_1(G)) + \mu(G),$$

or $n_1|E(G)| + 2p - \deg_1(G) + \mu(G)$. Thus $\deg_1(G) = 2p + \mu(G)$. □

Lemma 4.2. *Suppose that $\tau(P)$ is asymmetric, and let S be the set of degree 1 vertices in an edge of $\tau'(P)$. Then $|S| \leq 2^{m_2}$. Define nonnegative values j' and k' such that $|S| = \binom{m_2}{0} + \cdots + \binom{m_2}{j'} + k'$ with either $0 \leq k' < \binom{m_2}{j'+1}$ or $k' = 0$ and $j' = m_2$. Then*

$$w(S) \leq \sum_{i=0}^{j'} (m_2 - i)\binom{m_2}{i} + k'(m_2 - j' - 1).$$

Proof. For all $v_1, v_2 \in S$, there is an automorphism of the underlying unlabeled hypergraph of $\tau'(P)$ that switches v_1 and v_2 and fixes all other vertices. Thus all vertices in S must have different labels in $\tau'(P)$, which implies that $|S| \leq 2^{m_2}$. The lemma follows from the fact that S contains at most $\binom{m_2}{i}$ vertices with a label of cardinality $m_2 - i$. □

Let $w_{|S|}$ denote the upper bound on $w(S)$ in Lemma 4.2. Now suppose that $\tau'(P)$ is asymmetric and G is a component of $\tau'(P)$. A *defect* in G is one of the following. A *defective vertex* is a vertex v with degree at least 2 and weight less than m_2, counted with multiplicity $d(v) = m_2 - w(v)$. A *defective* edge is an edge e with set S of degree 1 vertices with collective weight less than $w_{|S|}$, counted with multiplicity $d(e) = w_{|S|} - w(S)$. The number of defects in G is denoted by $d(G)$.

Lemma 4.3. *Let G be a connected component of $\tau'(P)$. If $\tau(P)$ is asymmetric, then*

$$w(G) \leq m_2 \deg_2^+(G) + \sum_{e \in E(G)} w_{\deg_1(e)} - d(G).$$

Proof. Every vertex has weight at most m_2, and a defective vertex v with degree at least 2 has weight $m_2 - d(v)$. The set of degree 1 vertices in an edge e has weight $w_{\deg_1(e)}$ if e is not defective, and otherwise weight $w_{\deg_1(e)} - d(e)$. The lemma follows by adding over all vertices. □

If G contains $n_1|E(G)|/2 + p$ vertices, then write

$$2p + \mu(G) = b\left\lfloor \frac{2p + \mu(G)}{|E(G)|} \right\rfloor + b'\left\lceil \frac{2p + \mu(G)}{|E(G)|} \right\rceil$$

with nonnegative $b + b' = |E(G)|$. The following lemma states that the weight of G is maximized when all edges have about the same number of degree 1 vertices.

Lemma 4.4. *With all quantities as above,*

$$w(G) \le m_2 n_1|E(G)|/2 - m_2 p - m_2\mu(G) + bw_{\left\lfloor \frac{2p+\mu(G)}{|E(G)|} \right\rfloor} + b'w_{\left\lceil \frac{2p+\mu(G)}{|E(G)|} \right\rceil} - d(G).$$

Proof. By Lemma 4.1, G has $n_1|E(G)|/2 - p - \mu(G)$ vertices of degree at least 2. Then by Lemma 4.3,

$$w(G) \le m_2 n_1|E(G)|/2 - m_2 p - m_2\mu(G) + \sum_{e\in E(G)} w_{\deg_1(e)} - d(G).$$

The expression w_y is concave in y, meaning that for all y, $w_y - w_{y-1} \ge w_{y+1} - w_y$. Thus, given a set of values $\{y_i\}$ such that $\sum_i y_i = 2p + \mu(G)$, $\sum_{i=1}^{t} w_{y_i}$ is maximal when b of the y_i are equal to $\left\lfloor \frac{2p+\mu(G)}{|E(G)|} \right\rfloor$ and b' of the y_i are $\left\lceil \frac{2p+\mu(G)}{|E(G)|} \right\rceil$. The lemma follows. □

We now look to maximize the weight of G by considering the total number of vertices of a given weight. In particular, G contains at most $|E(G)|\binom{m_2}{i}$ degree 1 vertices with weight $m_2 - i$, since each each edge contains at most $\binom{m_2}{i}$ such vertices. Choose values j^* and k^* such that

$$2p + \mu(G) = |E(G)|\binom{m_2}{0} + \cdots + |E(G)|\binom{m_2}{j^*} + k^*$$

with $j^* \ge -1$ and $0 \le k^* < |E(G)|\binom{m_2}{j^*+1}$.

Lemma 4.5. *With all quantities as above,* $w(G) \le$

$$m_2\left(\frac{n_1|E(G)|}{2} - p - \mu(G)\right) + \sum_{i=0}^{j^*} |E(G)|(m_2 - i)\binom{m_2}{i} + (m_2 - j^* - 1)k^* - d(G).$$

Proof. Let G' be a (not necessarily asymmetric) hypergraph constructed from G by giving every vertex of degree at least 2 the label $[m_2]$ and assigning a label to every degree 1 vertex such that all edges of G' are nondefective and have distinct labels among the degree 1 vertices. Then G' has $(n_1|E(G)|/2 - p - \mu(G))$ vertices of degree at least 2, each of which has weight m_2, and at most $|E(G)|\binom{m_2}{i}$ degree 1 vertices of weight $m_2 - i$ for $0 \le i \le j^*$. The result follows by $w(G') = w(G) + d(G)$. □

We consider the upper bound of Lemma 4.5 to be a function $w_{\max}(p, \mu(G), d)$, with the quantities m_2 and $|E(G)|$ considered to be fixed. We note that

$$w_{\max}(p, \mu(G), d) > w_{\max}(p, \mu(G), d+1),$$

while

$$w_{\max}(p, \mu(G), d) \geq w_{\max}(p, \mu(G)+1, d),$$

with equality exactly when $j^* = -1$.

Now we consider $\mu(G) = d = 0$, and the upper bound of Lemma 4.5 is a function $w_{\max}(p)$. The effect of replacing p by $p+1$ is equivalent, numerically, to replacing a vertex with weight m_2 by two vertices, one of weight $m_2 - j^* - 1$ and the other of weight either $m_2 - j^* - 1$ or $m_2 - j^* - 2$. Thus the function $w_{\max}(p)$ is weakly unimodal in p and achieves a maximum when $j^* = \lfloor (m_2-1)/2 \rfloor$ and $k^* = 0$ or 1. Then $2p = \sum_{i=0}^{\lfloor (m_2-1)/2 \rfloor} |E(G)|\binom{m_2}{i} + (0 \text{ or } 1)$, and we have by Lemma 4.5 that

$$w(G) \leq |E(G)| \left(\frac{m_2 n_1}{2} - \frac{m_2}{2} \sum_{i=0}^{\lfloor \frac{m_2-1}{2} \rfloor} \binom{m_2}{i} + \sum_{i=0}^{\lfloor \frac{m_2-1}{2} \rfloor} (m_2 - i)\binom{m_2}{i} \right). \tag{4.1}$$

Thus the following holds.

Corollary 4.6. *Let all quantities be as above.*

1. *If* $j = \lfloor (m_2-1)/2 \rfloor$, *then* $w(G) \leq m_2 r |E(G)|$.
2. *If* $j < \lfloor (m_2-1)/2 \rfloor$ *and* $p \leq |E(G)|(\frac{n_1}{2} - 1)$, *then* $w(G) \leq m_2 r |E(G)|$.
3. G *has positive value only if* $j < \lfloor (m_2-1)/2 \rfloor$ *and* G *is a tree. Then if* G *has* d *defects,* $v(G) \leq m_2 - 2j - d$.

Proof. Part 1 follows by Equation (4.1) and the definition of r. Part 2 follows from the monotonicity of $w_{\max}$ and the definition of r. Part 3 is a consequence of Parts 1 and 2. □

We now focus on the particular case that G is a tree and $k = 0$. Suppose that G has l leaves, and by Lemma 2.1, $l \geq \mu(G) + 2$. Then $j^* = j$ and $k^* = 2 + \mu(G)$, and

$$\sum_{e \in E(G)} w_{\deg_1(e)} \leq \sum_{i=0}^{j} |E(G)|(m_2 - i)\binom{m_2}{i} + (\mu(G)+2)(m_2 - j - 1).$$

This bound can be attained if G contains $|E(G)|\binom{m_2}{i}$ degree 1 vertices of weight $m_2 - i$ for $0 \leq i \leq j$ and $\mu(G)+2$ degree 1 vertices of weight $m_2 - j - 1$. However, every leaf of G contains a vertex of weight at most $m_2 - j - 1$, and thus in fact $\sum_{e \in E(G)} w_{\deg_1(e)} \leq$

$$\sum_{i=0}^{j} |E(G)|(m_2 - i)\binom{m_2}{i} + (\mu(G)+2)(m_2 - j - 1) - (l - \mu(G) - 2).$$

Since G has $|E(G)| - 1 - \mu(G)$ vertices of degree 2 or more,

$$w(G) \leq m_2|E(G)| + \sum_{i=0}^{j} |E(G)|(m_2 - i)\binom{m_2}{i} - \mu(G)j + m_2 - 2j - l.$$

Finally, if we allow that G might have d defects, then we conclude the following.

Lemma 4.7. *With all quantities as above, $v(G) \leq m_2 - 2j - l - j\mu(G) - d$.*

Now we consider all of $\tau'(P)$. Let $\text{Comp}(P)$ be the set of connected components of $\tau'(P)$.

Lemma 4.8. *If $\tau(P)$ is asymmetric, $v(\tau'(P)) \leq \sum_{G \in \text{Comp}(P)} v(G) + m_2 2^{m_2 - 1}$.*

Proof. We calculate that $v(\tau'(P))$ is $\sum_{G \in \text{Comp}(P)} v(G)$ plus the sum of the weights of all vertices not contained in any n_1-edge. Since $\tau(P)$ is asymmetric, every vertex not contained in an n_1-edge must have a different label, and there is at most one vertex with every label $S \subset 2^{[m_2]}$. The lemma follows. □

5 An extremal construction

In this section, we give a general method of constructing a distinguishing partition P of $K_{m_1(n_1), m_2(n_2)}$. We then show that n_2 is maximal to within an additive constant, given the other parameters. We do so by describing the vertex-labeled hypergraph $\tau'(P)$. For the remainder of this section, we assume that $j < \lfloor (m_2 - 1)/2 \rfloor$; the case that $j = \lfloor (m_2 - 1)/2 \rfloor$ is treated seperately.

Let G be a component of $\tau'(P)$. Define the *value* $v(e)$ of an edge $e \in E(G)$ to be $v(G)/|E(G)|$. Let ξ be the map that adds 1 mod m_2 to every element in the label of every vertex of a $2^{[m_2]}$-vertex labeled hypergraph. Note that $v(\xi(G)) = v(G)$. Say that vertex labeled hypergraphs G, G' are equivalent under $\sim_\xi$ if $G = \xi^i(G')$ for some i.

Let $\mathcal{T}^* = \mathcal{T}^*_{n_1, m_2} = (\mathcal{T}_1, \mathcal{T}_2, \ldots)$ be an ordered list of equivalence classes under $\sim_\xi$ of positive weight asymmetric hypergraphs such that the edges in an element of $\mathcal{T}_i$ have value at least as great as the edges in an element of $\mathcal{T}_{i+1}$ for all i. By Lemma 4.7, an edge e may have value $\delta > 0$ only if e is contained in a tree with at most m_2/δ edges. Thus $\mathcal{T}^*$ enumerates all hypergraphs with positive value, and only classes of trees are in $\mathcal{T}^*$.

A *symmetry breaking loop* R is a $2^{[m_2]}$-vertex labeled hypergraph on at least $\min_R = \min_R(n_1, m_2)$ edges defined as follows. If $m_2 = 1$, then $j = -1$ and $n_1 = 2$, and we set $\min_R = 6$. Let R be a cycle that contains consecutive vertices v_1, v_2, v_3, v_4. Then all vertices of R have weight 1 except for v_1, v_2, v_4, which all have weight 0.

If $m_2 > 1$, we set $\min_R = \max(2m_2, n_1 + 1)$. Let quot_R be the maximum multiple of m_2 up to $|E(R)|$. Let $v_0, \ldots, v_{\text{quot}_R - 1}$ be vertices and $e_0, \ldots, e_{\text{quot}_R - 1}$ be edges such that e_i contains only degree 1 vertices except for v_i, v_{i+1}, subscripts mod quot_R. The set of vertex labels of the degree 1 vertices of e_0 includes all possible labels of size at least $m_2 - j$, and all others are of size $m_2 - j - 1$. To determine the labels of the degree 1 vertices of e_i, add i mod m_2 to every element in the labels of the degree 1 vertices of e_0. Assign v_0 the label $\emptyset$, v_i the label $[m_2] - i$ for $1 \leq i \leq m_2$, and v_i the label $[m_2]$ for all other i. Finally, R contains edges of the form $v_i, v_{i+1}, \ldots, v_{i+n_1-1}$ for $1 \leq i \leq |E(R)| - \text{quot}_R$.

We need some key facts on symmetry breaking loops.

Lemma 5.1. *Let R be a symmetry breaking loop that is a component of $\tau'(P)$. Then no automorphism of P induces a nontrivial automorphism of R.*

Proof. The lemma is readily verified when $m_2 = 1$, and so we assume that $m_2 > 1$. Let σ be an automorphism of $\tau(P)$ that induces an automorphism of R. Since v_0 is the only vertex of R of weight 0, it is a fixed point. Since v_1 is the only vertex of R that is in a common edge with v_0, has degree 2 in $\tau'(P)$, and weight not equal to m_2, v_1 is also a fixed point. Thus all v_i are fixed, which implies that σ fixes the n_2-edges of $\tau(P)$. Since all v_i are fixed, σ thus also fixes all n_1-edges of R. Finally, since all degree 1 vertices in a given edge have different labels, they must be fixed points as well. □

The following is readily observed from the construction of symmetry breaking loops.

Lemma 5.2. *Let R be a symmetry breaking loop, and let $1 \le i < i' \le m_2$. Then the number of vertices of R whose label contains i is equal to the number of vertices of R whose label contains i'.*

Lemma 5.3. *There exists a value $\omega = \omega_{n_1,m_2}$, which depends only on n_1 and m_2, such that a symmetry breaking loop R has value at least ω.*

Proof. If $m_2 = 1$, then $v(R) = -3$. If $m_2 > 1$, then the total weight of the degree 1 vertices in each edge with degree 1 vertices is $m_2(r-1)$. All other vertices have weight m_2 except for m_2 vertices of weight $m_2 - 1$ and one of weight 0. It follows that $w(R) = \text{quot}_R m_2 r - 2m_2$ and $v(R) = -(|E(R)| - \text{quot}_R)m_2 r - 2m_2 > -m_2^2 r - 2m_2$. □

We now come to our construction of P. Choose ζ to be the maximum value such that the total number of edges in all trees of $\mathcal{T}_1 \cup \cdots \cup \mathcal{T}_\zeta$ is at most $m_1 - \min_R$. Let $\Delta_{m_1(n_1),m_2}$ be the union of all trees in $\mathcal{T}_1 \cup \cdots \cup \mathcal{T}_\zeta$, together with a symmetry breaking loop so that $\Delta_{m_1(n_1),m_2}$ has m_1 edges, and a degree 0 vertex of every label except $\emptyset$.

Lemma 5.4. *$\Delta_{m_1(n_1),m_2}$ is in fact $\tau'(P)$ for a distinguishing partition P for an appropriate value of n_2.*

Proof. By construction, $\Delta_{m_1(n_1),m_2}$ contains the same number of vertices whose label contains i as the number of vertices whose label contains i' for all $1 \le i < i' \le m_2$. Thus $\Delta_{m_1(n_1),m_2}$ is $\tau'(P)$ for some partition P of $K_{m_1(n_1),m_2(n_2)}$ and for some n_2. Next, we apply Corollary 3.3 and show that P is distinguishing by showing that $\tau(P)$ is asymmetric. Let σ be an automorphism of $\tau(P)$. Since $\Delta_{m_1(n_1),m_2}$ contains exactly one symmetry breaking loop, all n_2-edges of $\tau(P)$ are fixed under σ. Since all components of $\Delta_{m_1(n_1),m_2}$ are asymmetric and no two are isomorphic to each other, all n_1-edges and vertices of $\tau(P)$ are fixed as well. □

Next, we prove that $\Delta_{m_1(n_1),m_2}$ is a nearly optimal construction.

Lemma 5.5. *Let P be a distinguishing partition of $K_{m_1(n_1),m_2(n_2)}$ such that $\tau'(P) = \Delta_{m_1(n_1),m_2}$, and let G' be an asymmetric vertex-labeled hypergraph. Then $w(G') \le w(\Delta_{m_1(n_1),m_2}) + \text{Error}_{n_1,m_2}$ for some value Error_{n_1,m_2} that depends only on n_1 and m_2. In particular, if P' is a distinguishing partition of $K_{m_1(n_1),m_2(n_2')}$, then $n_2' \le n_2 + \text{Error}_{n_1,m_2}/m_2$.*

Proof. First we determine an upper bound on $w(G')$ in terms of the structure $\mathcal{T}^*$, and then we compare that to $w(\Delta_{m_1(n_1),m_2})$. Let v^+ be the sum of the weights of all vertices of G' that are not contained in edges; since they must all have different labels, $v^+ \le m_2 2^{m_2-1}$. Then, summing over all components G of G' that contain n_1-edges, $n_2' \le \frac{1}{m_2}\sum_G w(G) + 2^{m_2-1} = rm_1 + \frac{1}{m_2}\sum_{e \in E(G')} v(e) + 2^{m_2-1}$.

Suppose that the set of edge values of hypergraphs in $\mathcal{T}^*$ are $\{v_1, v_2, \ldots\}$ with $v_1 > v_2 > \cdots$, and suppose that there are nv_i total edges in all hypergraphs of $\mathcal{T}^*$ with value v_i. Let ρ be the largest value such that $\tau'(P)$ contains nv_ρ edges of value v_ρ. Then

$$w(G') \le rm_1m_2 + \sum_{i=1}^{\rho} \mathrm{nv}_i v_i + \left(m_1 - \sum_{i=1}^{\rho} \mathrm{nv}_i\right) v_{\rho+1} + m_2 2^{m_2-1}.$$

Now we consider $\tau'(P)$ with symmetry breaking loop R. Choose ζ so that $\mathcal{T}_\zeta$ is the last equivalence class of trees that are components of $\tau'(P)$; edges in $\mathcal{T}_{\zeta+1}$ have value $v_{\rho+1}$, and thus by Lemma 4.7, each tree in $\mathcal{T}_{\rho+1}$ has at most $m_2/(v_{\rho+1})$ edges. By construction, R has at most $m_2^2/(v_{\rho+1}) + \min_R$ edges, each of which has value at least $\omega/|E(R)|$. Furthermore, $\tau'(P)$ contains nv_i edges of value v_i for $1 \le i \le \rho$, and all edges besides these and edges in R have value $v_{\rho+1}$. Thus $w(\Delta_{m_1(n_1),m_2}) \ge$

$$rm_1m_2 + \sum_{i=1}^{\rho} \mathrm{nv}_i v_i + \left(m_1 - \sum_{i=1}^{\rho} \mathrm{nv}_i\right) v_{\rho+1} - |E(R)| \left(v_{\rho+1} - \frac{\omega}{|E(R)|}\right) + m_2 2^{m_2-1}.$$

Thus $w(G') - w(\Delta_{m_1(n_1),m_2}) \le |E(R)|(v_{\rho+1} - \omega/|E(R)|) \le m_2^2 + \min_R v_{\rho+1} - \omega$. This proves the lemma. □

In the subsequent sections, we prove upper bounds on n_2 in terms of the other variables by evaluating the weights of $\Delta_{m_1(n_1),m_2}$. For every m_1, n_1, m_2, choose $n_2'(m_1, n_1, m_2)$ maximally so that $K_{m_1(n_1),m_2(n_2')}$ has a distinguishing partition $P'_{m_1(n_1),m_2}$.

Lemma 5.6. $\lim_{m_1\to\infty} \frac{n_2'(m_1,n_1,m_2)}{n_2'(m_1+1,n_1,m_2)} = 1.$

Proof. It follows by construction of $\Delta_{m_1(n_1),m_2}$ and Lemma 5.5. □

Lemma 5.7. *If m_1 is sufficiently large relative to m_2, n_1, and n_2, then $K_{m_1(n_1),m_2(n_2)}$ has a distinguishing partition.*

Proof. If n_2 is small relative to n_1 and m_2, then an asymmetric hypergraph with m_i edges of size n_i for $i = 1, 2$ may be constructed as follows. First take an asymmetric hypergraph with m_1 n_1-edges, which exists by the main result of [7]. Then add m_2 n_2-edges on the same vertex set. The result is asymmetric. □

Now assume that n_2 is large. Choose m^* maximally so that $n_2' = n_2'(m^*, n_1, m_2) \ge n_2$. By Lemma 5.6, n_2'/n_2 is close to 1. We need to show that $K_{m^*(n_1),m_2(n_2)}$ has a distinguishing partition. Our method is to show that there exists distinguishing partition P' of $K_{m^*(n_1),m_2(n_2')}$ and a subset S of weight m_2 vertices on $\tau'(P')$, with $|S| = n_2' - n_2$, such that the hypergraph that results by changing all of the labels of vertices of S from $[m_2]$ to $\emptyset$ is asymmetric.

Lemma 5.8. *With all quantities as above, $V(\tau'(P'))$ has a subset S of weight m_2 vertices of size $n_2' - n_2$ such that the hypergraph G' that results from changing all labels of vertices of S from $[m_2]$ to $\emptyset$ is $\tau'(P^*)$ for a distinguishing partition P^* of $K_{m^*(n_1),m_2(n_2)}$.*

Proof. Certainly P^* is a partition of $K_{m^*(n_1),m_2(n_2)}$. It suffices to show that S may be chosen so that $\tau(P^*)$ is asymmetric. If $\tau'(P')$ contains $\Theta(n_2')$ vertices of degree at least 2 of weight m_2 and none of weight 0, then any set S of size $n_2' - n_2$ of weight m_2 and degree at least 2 vertices satisfies the desired property. This condition is seen directly in all cases that $j = \lfloor (m_2 - 1)/2 \rfloor$, as P' is constructed directly in subsequent sections.

It must be that $\tau'(P')$ has few components with weight 0 vertices, and no components with many weight 0 vertices. Otherwise, if $n_1 = 2$, we could replace all those components and a largest defect-free component of $\tau'(P')$, if there is one, by a single defect-free asymmetric tree, which would increase the weight, a contradiction to Lemma 5.5. Otherwise, we could replace all those components and a largest symmetry breaking loop of $\tau'(P')$, if there is one, with a single symmetry breaking loop. By Lemma 4.7 if $n_1 = 3$, and otherwise by Corollary 4.6, this would increase the weight, also a contradiction to Lemma 5.5.

It is shown in subsequent sections that for all sufficiently large t, there is an element of $\mathcal{T}^*$ with t edges. Thus there are $\Omega(\sqrt{m^*})$ elements of $\mathcal{T}^*$ of size up to $2\sqrt{m_1}$ that are not components of $\tau'(P')$. It must be that all but $o(m^*)$ edges of $\tau'(P')$ are contained in positive weight components; otherwise, all components with nonpositive weight could be removed and replaced by $\Omega(\sqrt{m^*})$ components of positive weight, a contradiction to Lemma 5.5.

Let $T_1, \dots, T_a$ be the components of $\tau'(P')$ with weight 0 vertices. Let T be a positive weight component with t vertices. Then T has $\Theta(t)$ vertices of degree at least 2, all but at most m_2 of which have weight m_2, and thus $\tau'(P')$ has $\Theta(m^*)$ vertices in positive weight components with weight m_2. Let S be a subset of size $n_2' - n_2$, chosen uniformly at random, of the weight m_2 vertices that are contained in the larger half of positive weight components. Let G' be the hypergraph that results from changing the labels of all vertices of S in $\tau'(P')$ from $[m_2]$ to $\emptyset$.

Every component T of G' that contains a vertex of S is asymmetric, since it was constructed by dividing the set of vertices labeled $[m_2]$ into vertices labeled $[m_2]$ and $\emptyset$, and it had no vertex labeled $\emptyset$ previously. T is not isomorphic to another component T' that contains a vertex of S since the hypergraph that results from changing all vertices of T of label $\emptyset$ to $[m_2]$ is nonisomorphic to the hypergraph that results from changing all vertices of T' of label $\emptyset$ to $[m_2]$. To conclude, we need to show that with high probability, T_i is not isomorphic to any component of G' for each $1 \le i \le a$, since a is small.

If T_i is isomorphic to T, a component of G' that contains a vertex of S, then it must be that the hypergraphs T_i^* and T^*, which result from converting all vertices of T_i and T of label $\emptyset$ to $[m_2]$, are isomorphic. Thus T_i^* is isomorphic to a component T^* of $\tau'(P')$. Since T^* is asymmetric, for all subsets S' of weight m_2 vertices of T^*, the hypergraphs that result from converting all vertices of S' from label $[m_2]$ to $\emptyset$ are nonisomorphic. Since T^* is large, the probability that T_i is isomorphic to some other component of G' is small. □

Lemma 5.9. *Suppose that $K_{m^*(n_1),m_2(n_2)}$ has a distinguishing partition and $m_1 > m^*$. Then $K_{m_1(n_1),m_2(n_2)}$ has a distinguishing partition.*

Proof. It suffices to prove the lemma for $m_1 = m^* + 1$. Let P be a distinguishing partition of $K_{m^*(n_1),m_2(n_2)}$. Define a *tail T of $\tau(P)$* to be a sequence of vertices $v_1, \dots, v_{n_1+t-1}$ and edges $\{v_i, \dots, v_{n_1+i-1}\}$ for $1 \le i \le t$, such that $v_{n_1}, \dots, v_{n_1+t-1}$ are contained in no edges outside of T. Assume that T is chosen so that t is maximal. Then add an vertex v_{n_1+t} and an edge $\{v_{t+1}, \dots, v_{n_1+t}\}$ to $\tau(P)$ to create a hypergraph G'.

We show that G' is asymmetric. By construction, v_{n_1+t} is the only degree 1 vertex contained in a maximum tail of G', and thus it is a fixed point. Then the edge $\{v_{t+1}, \ldots, v_{n_1+t}\}$ is fixed, since it is the only edge to contain v_{n_1+t}. Thus all other vertices and edges are fixed as well, since $\tau(P)$ is asymmetric. It follows that $K_{m_1(n_1),m_2(n_2)}$ has a distinguishing partition. □

We summarize the preceding lemmas as follows.

Corollary 5.10. *If m_1 is large relative to n_1 and m_2, let n_2' be the largest value such that $K_{m_1(n_1),m_2(n_2')}$ has a distinguishing partition. Then $K_{m_1(n_1),m_2(n_2)}$ has a distinguishing partition if $n_2 \leq n_2'$.*

6 $n_1 = 2$

In this section we consider the case that $n_1 = 2$. A labeled connected 2-uniform hypergraph has positive value only if it is a tree with fewer than m_2 defects. Furthermore, all trees without defects have positive value.

Our bounds and construction requires an estimate on the number of asymmetric ordinary trees, which is provided by the twenty-step algorithm of Harary, Robinson, and Schwenk.

Lemma 6.1. *There exists constants $\alpha > 0$ and $\beta > 1$ such that the number of asymmetric trees on i edges is $(1 + o_i(1))\alpha\beta^i i^{-5/2}$. Furthermore, there exists $\alpha' > 0$ such that the number of asymmetric rooted trees on i edges is $(1 + o_i(1))\alpha'\beta^i i^{-3/2}$.*

Proof of Theorem 1.2: Recall that

$$z = \left\lfloor \log_\beta \left(\frac{m_1(\beta - 1)}{\alpha\beta} \left(\log_\beta m_1\right)^{3/2} \right) \right\rfloor .$$

Summing the result of Lemma 6.1 from 1 to z, the number of nonisomorphic asymmetric trees with at most z edges is $(\alpha\frac{\beta}{\beta-1} + o(1))\beta^z z^{-5/2}$, and they collectively have $(\alpha\frac{\beta}{\beta-1} + o(1))\beta^z z^{-3/2} \leq (1 + o(1))m_1$ edges. Similarly, the collective number of edges of nonisomorphic asymmetric trees with at most $z + 1$ edges is at least $(1 + o(1))m_1$, and with at most $z + 2$ edges exceeds m_1. Each edge of a defect-free tree on $z + 2$ edges has value $m_2/(z + 2)$, and thus all edges of $\Delta_{m_1(n_1),m_2}$ have value at least $m_2/(z+2)$, except edges in the symmetry breaking loop.

We show that $\Delta_{m_1(n_1),m_2}$ has $\frac{m_1}{z+1} + (1 + o_{m_1}(1))\alpha\beta^z z^{-7/2} \left(\frac{\beta}{\beta-1}\right)^2$ components of value m_2 and $o(\beta^z z^{-7/2}) = o(m_2/z^2)$ components of lesser value. The latter statement follows by Lemma 6.2. Thus in fact $\Delta_{m_1(n_1),m_2}$ contains $(1+o_i(1))\alpha'\beta^i i^{-3/2}$ defect-free components on i edges for $1 \leq i \leq z$, $o(m_1/z)$ defect-free components on $z + 2$ edges, $o(m_1/z^2)$ components of other types, and all remaining components are defect-free on $z + 1$ edges.

For every edge $e \in E(\Delta_{m_1(n_1),m_2})$, let $v^*(e) = v(e) - m_2/(z + 1)$. For $1 \leq i \leq z$, the sum of $v^*(e)$ over all edges e in components with $z + 1 - i$ edges is thus

$$(m_2 + o(1))\alpha\beta^{z+1-i}(z + 1 - i)^{-3/2}(1/(z + 1 - i) - 1/(z + 1)).$$

By $\beta^z z^{-7/2} = \Theta(m_1/\log^2 m_1)$, the preceding sum is

$$m_2\alpha\beta^{z+1-i}z^{-3/2}(i/z^2) + o(m_1\beta^{-i}/\log^2 m_1)$$

for $i < z/\log z$, and otherwise

$$m_2\alpha\beta^{z+1-i}z^{-3/2}(i/z^2) + o(m_1/\log^3 m_1),$$

which is observed by noting that $\beta^{z-z/\log z} = O((m_2\log(m_1)^{3/2})^{1-1/\log z})$, and that $m_1^{1/\log z} > m_1^{i\log\log m_1/\log m_1} = \log^i(m_1)$ for all fixed i.

The sum of $v^*(e)$ over all edges e in components with either $z+2$ edges or with defects is $o(m_1/\log^2(m_1))$, whereas $v^*(e) = 0$ if e is in a defect-free component with $z+1$ edges. We conclude that

$$\sum_{e\in E(\tau'(P))} v^*(e) = \alpha\beta^z z^{-7/2}\sum_{i=0}^{z}\beta^{-i}(i+1) + o(m_1/\log^2 m_1)$$

$$= \alpha\beta^z z^{-7/2}\frac{\beta^2}{(\beta-1)^2} + o(m_1/\log^2 m_1).$$

Thus

$$\sum_{e\in E(\tau'(P))} v(e) = \frac{m_1}{z+1} + \alpha\beta^z z^{-7/2}\frac{\beta^2}{(\beta-1)^2} + o(m_1/\log^2 m_1),$$

which implies that G has the desired number of components of value m_2.

The proof of Lemma 6.2 makes use of the species $L(\mathfrak{a}^\bullet)$ of ordered sets of rooted asymmetric trees: an element of $L(\mathfrak{a}^\bullet)$ on z' elements is given by order partitioning $[z']$ into subsets and taking an $\mathfrak{a}^\bullet$-structure on each subset.

Lemma 6.2. *There are $o(m_1/z^2)$ components of $\Delta_{m_1(n_1),m_2}$ with defects.*

Proof. If G is a component of $\Delta_{m_1(n_1),m_2}$ with a defect, then the value of G is at most $m_2 - 1$, and thus since the edges of G have value at least $m_2/(z+2)$, then G has at most $\frac{m_2-1}{m_2}(z+2)$ edges. Thus we need to show that the number of components with defects on at most $\frac{m_2-1}{m_2}(z+2)$ edges in $\Delta_{m_1(n_1),m_2}$ is $o(\beta^z z^{-9/2})$. It suffices to show that the number of components with positive value on z' edges is $O((\beta')^{z'})$ for fixed $\beta < \beta' < \beta^{m_2/(m_2-1)}$ and $z' < \frac{m_2-1}{m_2}(z+2)$, since $\beta^{z'} < \beta^z/z^i$ for fixed i.

If G is a component of positive value with z' edges, then by Lemma 4.7 all but at most $m_2 - 1$ vertices of G are labeled $[m_2]$. Let G' be the subgraph of G that is the union of all paths between vertices not labeled $[m_2]$. Only vertices not labeled $[m_2]$ are leaves in G', and each leaf has one of $2^{m_2} - 1$ labels, and thus the number of such G' that may result is at most a polynomial in z', say $p(z')$. We may reconstruct G from G' by replacing every vertex $v \in G'$ with a rooted asymmetric tree with root v. Thus, since G is determined by G' and an ordered set of asymmetric trees on a total of $z'+1$ vertices, there are at most $p(z')|L(\mathfrak{a}^\bullet)_{z'+1}|$ components on z' edges.

Choose fixed $\beta < \beta^* < \beta'$. The lemma follows by showing that $|L(\mathfrak{a}^\bullet)_{z'+1}| = O((\beta^*)^{z'})$. We show inductively on z' that $|L(\mathfrak{a}^\bullet)_{z'+1}| \le \gamma(\beta^*)^{z'}$ for some sufficiently large γ.

Note the recursion $L(\mathfrak{a}^\bullet) = E_0 + \mathfrak{a}^\bullet L(\mathfrak{a}^\bullet)$: every ordered set of rooted trees is either the empty set or a rooted tree followed by another ordered set of rooted trees. Thus for $z' \geq 1$,

$$|L(\mathfrak{a}^\bullet)_{z'+1}| = \sum_{i=1}^{z'+1} |\mathfrak{a}_i^\bullet||L(\mathfrak{a}^\bullet)_{z'+1-i}| \leq \sum_{i=1}^{z'+1} |\mathfrak{a}_i^\bullet|\gamma(\beta^*)^{z'+1-i}.$$

Let $\mathfrak{a}^\bullet_{\leq z'/3}$ be the species of rooted asymmetric trees on at most $z'/3$ vertices. Observe that $|(\mathfrak{a}^\bullet_{\leq z'/3}\mathfrak{a}^\bullet)_{z'+1}| \leq |\mathfrak{a}^\bullet_{z'+1}|$: given an asymmetric rooted tree T on $i \leq z'/3$ vertices and another T' on $z' + 1 - i$ vertices, a third tree may be constructed by adjoining T' to T so that the root of T' is forgotten and place adjacent to the root of T. This construction allows T and T' to be uniquely determined. Thus by Lemma 6.1,

$$\sum_{i=1}^{\lfloor z'/3 \rfloor} |\mathfrak{a}_i^\bullet||\mathfrak{a}^\bullet_{z'+1-i}| \leq (1 + o(1))\alpha'\beta\beta^{z'} z'^{-3/2}.$$

But also,

$$\sum_{i=1}^{\lfloor z'/3 \rfloor} |\mathfrak{a}_i^\bullet||\mathfrak{a}^\bullet_{z'+1-i}| \geq \alpha'\beta\beta^{z'} z'^{-3/2}(1 + o(1)) \sum_{i=1}^{\lfloor z'/3 \rfloor} |\mathfrak{a}_i^\bullet|\beta^{-i}.$$

Thus

$$\sum_{i=1}^{\lfloor z'/3 \rfloor} |\mathfrak{a}_i^\bullet|\beta^{-i} \leq 1 + o(1) \quad \text{and} \quad \sum_{i=1}^{\lfloor z'/3 \rfloor} |\mathfrak{a}_i^\bullet|(\beta^*)^{-i} < \beta/\beta^*.$$

Thus $\sum_{i=1}^{z'+1} |\mathfrak{a}_i^\bullet|\gamma(\beta^*)^{z'+1-i} \leq$

$$\gamma\beta(\beta^*)^{z'} + \sum_{i=\lfloor z'/3 \rfloor+1}^{z'+1} |\mathfrak{a}_i|\gamma(\beta^*)^{z'+1-i} =$$

$$\gamma\beta(\beta^*)^{z'} + (1 + o(1)) \sum_{i=\lfloor z'/3 \rfloor+1}^{z'+1} \alpha'\beta^i i^{-3/2}\gamma(\beta^*)^{z'+1-i} < \gamma(\beta^*)^{z'+1}$$

as desired. □

7 $k = 0$ and $j < \lfloor(m_2 - 1)/2\rfloor$

In this section we consider the case that $k = 0$ and $j < \lfloor(m_2 - 1)/2\rfloor$. The value of C defined in the statement of Theorem 1.3 is determined as follows. If $0 < j < m_2/2 - 3/2$, then

$$C = \binom{m_2}{j+1}^{m_2-2j-1} \binom{m_2}{j}^{m_2-2j-3} \binom{2m_2 - 4j - 4}{m_2 - 2j - 3} \frac{2^{-m_2+2j+3}}{(2m_2 - 4j - 4)!}.$$

If $j = 0$ and $m_2 > 3$, then

$$C = m_2^{m_2-1} \binom{2m_2 - 4}{m_2 - 3} \frac{1}{(2m_2 - 4)!},$$

and if $j = m_2/2 - 3/2$, then $C = \binom{m_2}{j+1}\left(\binom{m_2}{j+1} - 1\right)/2$.

The result requires the following estimate on the number of structures of a particular type of tree.

Lemma 7.1. *If i is odd, the number of labeled structures of $\mathfrak{A}_{\mathcal{E}_1+\mathcal{E}_3}$ on $i+1$ vertices is $\binom{i+1}{(i-1)/2}\frac{(i-1)!}{2^{(i-1)/2}}$, and of $\mathfrak{A}_{\mathcal{E}_1+2\mathcal{E}_3}$ on $i+1$ vertices is $\binom{i+1}{(i-1)/2}(i-1)!$.*

Proof. Apply Proposition 3.1.19 of Bergeron, Labelle, and Leroux [4]. □

Proof of Theorem 1.3: We start by proving an upper bound on the sum of the values of all components of $\Delta_{m_1(n_1),m_2}$. Let G be a component of $\Delta_{m_1(n_1),m_2}$ with t edges and positive value; by Lemma 4.7, G is a tree with at most $m_2 - 2j - 1$ leaves.

Construct a colored graph $c(G)$ from G as follows: $V(c(G))$ is the union of all edges of G and vertices of G of degree at least two; and $E(c(G))$ is given by vertex-edge containment. Every defective vertex or edge in G is colored red in $c(G)$. If v is a non-defective vertex of degree at least 3, then v is colored green in $c(G)$. All other vertices of $c(G)$ are blue.

A *segment* in $c(G)$ is a maximal path $(v_0, \ldots, v_i)$ such that for all $0 < i' < i$, $v_{i'}$ is a blue vertex with degree 2. Construct a new graph $c'(G)$ by replacing every segment $(v_0, \ldots, v_i)$ with a single edge v_0v_i, and label that edge by the number i of edges it replaces in $c(G)$.

The number of edges of $c'(G)$ is at most $2m_2 - 4j - 5$. To see this, observe that G has d defects and at most $m_2 - 2j - 1 - d$ leaves by Lemma 4.7. Every leaf of $c'(G)$ is a leaf of G. Combining the facts that $\sum_{v\in V(c'(G))}\deg(v) = 2e(c'(G))$ and $\sum_{v\in V(c'(G))}(\deg(v) - 2) = -2$, $e(c'(G)) \leq 2m_2 - 4j - 5 - 2d + a - b$, where a and b are the number of vertices of degree 2 and at least 4 in $c'(G)$. However, the only degree 2 vertices in $c'(G)$ correspond to defects in G, and thus $a \leq d$. Thus $c'(G)$ has at most $2m_2 - 4j - 5$ edges. Furthermore, this bound is attained only if G has $m_2 - 2j - 1$ leaves, no defects, no vertices of degree at least 4, and no vertices of degree 3 if $j > 0$ by Lemma 4.7.

If $c'(G)$ has $2m_2 - 4j - 5$ edges, then G has value 1 and no edges that intersects four other edges, since this would give a vertex of degree at least 4 in $c'(G)$. Thus, if $c'(G)$ has $2m_2 - 4j - 5$ edges, then $c'(G)$ has $m_2 - 2j - 1$ leaves and $m_2 - 2j - 3$ vertices of degree 3. If $j > 0$, then by Lemma 4.7, all vertices of $c'(G)$ are blue, and such trees, forgetting labels, may be described by the species $\mathfrak{A}_{\mathcal{E}_1+\mathcal{E}_3}$. If $j = 0$, then the degree 3 vertices may be green or blue, and thus such trees are described by the species $\mathfrak{A}_{\mathcal{E}_1+2\mathcal{E}_3}$.

Given $c'(G)$ with $i = 2m_2 - 4j - 5 \geq 2$ edges and that G has t edges, there are $(1+o_1(t))t^{i-1}/((i-1)!\alpha)$ nonisomorphic labellings of the edges, where α is the cardinality of the automorphism group of $c'(G)$. This is since that in most labellings, all labels are distinct, and the orbit of a labeling with distinct labels consists of α labellings. The number of labeled graphs $c'(G)$ of a given isomorphism class and automorphism group of order α is $(i+1)!/\alpha$. Hence the number of graphs $c(G)$ with $c'(G)$ having $m_2 - 2j - 1$ leaves is $\gamma(1 + o_1(t))t^{i-1}/((i-1)!(i+1)!)$, where γ is the number of labeled specimens of $\mathfrak{A}_R$ as in Lemma 7.1.

If $c'(G)$ has fewer than $i = 2m_2 - 4j - 5$ edges, there are $o(t^{i-1})$ labeling of $c'(G)$. Since the number of graphs $c'(G)$ that may arise is independent of t, the total number of graphs $c(G)$ is $\gamma(1 + o_1(t))t^{i-1}/((i-1)!(i+1)!)$, of which almost all have $m_2 - 2j - 1$ leaves and all segments of different lengths.

Given a graph G', the number of components G of $\Delta_{m_1(n_1),m_2}$ with positive value such that $c(G) = G'$ has an upper bound that depends only on n_1 and m_2. G is determined by $c(G)$ and the following: the labels of all defective vertices, the labels of all vertices that are contained in defective edges, the labels of all vertices contained in leaves of G, and the labels of all vertices contained in edges with at least 3 vertices of degree at least 2. There are at most $m_2 - 2j - 1$ of each of these items in G, and each one may be determined in at most $2^{n_1 m_2}$ ways, and thus there are at most $2^{4n_1 m_2(m_2-2j-1)}$ components G with positive value such that $c(G) = G'$. Thus, there are $o(t^{2m_2-4j-6})$ positive-value components G with t edges such that either G has at most $m_2 - 2j - 2$ leaves or $c'(G)$ has two edges with the same label. Adding over all t, there are $o(m_1^{\frac{2m_2-4j-5}{2m_2-4j-4}})$ such components G.

Now we determine how many positive-value components G with t edges of $\Delta_{m_1(n_1),m_2}$ satisfy these two conditions: $c(G) = G'$ for a particular graph G' with $m_2 - 2j - 1$ leaves, and $c'(G)$ has distinct edge labels. If $0 \le j < m_2/2 - 3/2$, G has no defects, no vertices of degree at least 3 (if $j > 0$), and $m_2 - 2j - 3$ edges that intersect 3 others. Each leaf, since it is not defective, contains one vertex of every label S with $|S| \ge m_2 - j$ and exactly one vertex with a label S with $|S| = m_2 - j - 1$. There are $\binom{m_2}{j+1}$ ways to select this label. Each edges that intersects 3 others, since it is not defective, contains a vertex of every label S with $|S| \ge m_2 - j$ except for one label S with $|S| = m_2 - j$. There are $\binom{m_2}{j}$ ways to choose this label. The total number of such components is $\binom{m_2}{j+1}^{m_2-2j-1}\binom{m_2}{j}^{m_2-2j-3}$, and the distinct edge labels of $c'(G)$ ensure that each of these components are asymmetric.

If $j = m_2/2 - 3/2$, then $c(G)$ has 2 leaves and is a path. As before, the two leaves each contain a vertex of every label S with $|S| \ge m_2 - j$, together with one vertex each of labels S and S' respectively with $|S| = |S'| = m_2 - j - 1$. All other edges contain exactly a vertex of each label $\tilde{S}$ with $|\tilde{S}| \ge m_2 - j$. By asymmetry, $S \ne S'$. Thus there are $\binom{m_2}{j+1}(\binom{m_2}{j+1} - 1)/2$ asymmetric components G with $c(G) = G'$.

We conclude that there are $(C + o(1))t^{2m_2-4j-6}$ components of $\Delta_{m_1(n_1),m_2}$ with t edges and positive value, almost all of which have value 1 and none with value exceeding $m_2 - 2j - 2$. Adding over all

$$t < (1 + o(1)) \left(\frac{m_1(2m_2 - 4j - 4)}{C}\right)^{\frac{1}{2m_2-4j-4}}$$

proves the result.

8 $k \ge 1$ and $j < \lfloor (m_2 - 1)/2 \rfloor$

Proof of Theorem 1.4: We start with the upper bound on n_2. By Lemma 4.7, every component of $\Delta_{m_1(n_1),m_2}$ has value at most $m_2 - 1$, and every component with positive value is a tree. Suppose that the number of components of $\Delta_{m_1(n_1),m_2}$ on i edges with positive value is at most b^i for some b. Then $\Delta_{m_1(n_1),m_2}$ has at most $\frac{b^{\lceil(\log_b(m_1))/2\rceil+1}-1}{b-1}$ components with at most $\lceil \log_b(m_1)/2 \rceil$ edges. Thus, $\Delta_{m_1(n_1),m_2}$ has at most $\frac{b^{\lceil(\log_b(m_1))/2\rceil+1}-1}{b-1} + \frac{m_1}{\lceil \log_b(m_1)/2 \rceil}$ components of positive value, each of which has value at most $m_2 - 1$. This would prove the upper bound.

We now establish that there are at most b^i components of positive value on i edges for some b. Every component G of positive value is a tree. Associate with G a labeled tree G' as follows. The vertex set of G' is the union of the edge set of G and the set of vertices of G with degree at least 2. The edges of G' are given by inclusion in G. If v is a vertex of G' that corresponds to a vertex of G, then v is given the same label; thus, there are at most 2^{m_2} possible labels for v. If e is a vertex of G' that corresponds to an edge of G, then e is labeled in a way to encode the number and labels of degree 1 vertices of e. Thus e can be labeled in at most $1 + 2^{m_1} + 2^{2m_1} + \cdots + 2^{n_2 m_1}$ ways. G can be reconstructed to isomorphism from G'.

If G has i edges, then G' has at most $2i - 1$ vertices. Thus the number of isomorphism classes of underlying unlabeled trees of G' grows exponentially in i [10]. Since the number of possible labels of each vertex of G' depends only on n_1 and m_2, the total number of trees G', and thus components G, grows at most exponentially in i.

To prove the lower bound on n_2, we show that the number of components of $\Delta_{m_1(n_1),m_2}$ with t edges does in fact grow exponentially in t. Let G be a tree without defects or vertices of degree at least 3 such that every edge contains at least $\binom{m_2}{0} + \cdots + \binom{m_2}{j}$ degree 1 vertices. Then G contains $2|E(G)| - 1$ vertices of weight m_2, $|E(G)|\binom{m_2}{i}$ of weight $m_2 - i$ for $1 \leq i \leq j$, and $|E(G)|k + 2$ of weight $m_2 - j - 1$. Then G has value $m_2 - 2j - 2$.

To G we may associate a vertex-labeled tree G', with the vertices of G' given by the edges of G, and the edges of G' are given by intersection. The label of a vertex of G' encodes the labels of the degree 1 vertices of the corresponding edge. Suppose that such an edge e intersects i other edges. Since e contains a vertex of every label with of size at least $m_2 - j$ and $k - i + 2$ vertices of label of size $m_2 - j - 1$, there are $a_i := \binom{\binom{m_2}{j+1}}{k-i+2}$ ways to label e in G'. Thus G' may be regarded as a member of $\mathfrak{a}_{\sum_{i=1}^{k+2} a_i \mathcal{E}_i}$, and from every member of this species, one can reconstruct an asymmetric $2^{[m_2]}$-labeled n_1-uniform tree with positive value. We show that $|(\mathfrak{a}_{\sum_{i=1}^{k+2} a_i \mathcal{E}_i})_t|$ grows exponentially in t by exhibiting a subset of structures of exponential size.

Let $(v_0, \ldots, v_{\lfloor 2t/3 \rfloor})$ be a path. Let S be subset of size $t - 1 - \lfloor 2t/3 \rfloor$ of the integers from 3 to $\lfloor 2t/3 \rfloor - 2$ that includes 3 and $\lfloor 2t/3 \rfloor - 2$. For every $i \in S$, let u_i be a vertex with an edge $u_i v_i$. Then the graph with vertices $(v_0, \ldots, v_{\lfloor 2t/3 \rfloor})$ and u_i for each $i \in S$, and labels chosen arbitrarily, is an element of $\mathfrak{a}_{\sum_{i=1}^{k+2} a_i \mathcal{E}_i}$. Two such graphs are nonisomorphic for different choices of S, and the number of choices of S grows exponentially in t.

Say that there are b^t trees of the maximum possible value of $m_2 - 2j - 2$ on t edges for some fixed b and sufficiently large t. Then $\Delta_{m_1(n_1),m_2}$ contains no component with more than $\lceil \log_b(m_1) \rceil$ edges for large m_1, except possibly the symmetry breaking loop, and $\Delta_{m_1(n_1),m_2}$ has at least $\frac{m_1 - m_2\lceil \log_b(m_1)\rceil - \min_R}{\lceil \log_b(m_1)\rceil}$ components. This proves the theorem.

9 $k = 0$ and $j = \lfloor (m_2 - 1)/2 \rfloor$

We consider Theorem 1.5 in three cases.

Theorem 9.1. *Theorem 1.5 holds for even $m_2 \geq 4$.*

Proof. The upper bound on n_2 follows from Lemmas 4.7 and 4.8: when $j = (m_2 - 1)/2$, since every tree has at least 2 leaves, no component has positive value.

We establish the that n_2 may be $rm_1 + 2^{m_2 - 1}$ by the following construction. Let $\tau'(P)$ consist of components $G_1, \ldots, G_{m_2}$ of n_1-edges, such that G_i consists of t_i edges, and all the t_i are distinct and sum to m_1. Say that G_i contains edges $e_1, \ldots, e_{t_i}$ such that for all $1 \leq a < b \leq t_i$, e_a and e_b do not intersect unless $b = a + 1$, in which case $e_a \cap e_b = \{v_a\}$. Each v_a is labeled $[m_2]$. Each e_a contains one vertex of each of label of size at least $m_2 - j$.

In addition, e_1 and e_{t_i} contain respective vertices u_1 and u_2 of labels $\{i, i+1, \dots, i+j\}$ and $\{i, i-1, \dots, i-j\}$, subscripts mod m_2. In addition, $\tau'(P)$ contains one degree 0 vertex of each nonempty label.

Now we show that $\tau(P)$ is asymmetric. Since the G_i have different numbers of edges, no automorphism permutes the components of $\tau'(P)$ nontrivially. The only nontrivial automorphism of the edges of G_i reverses the chain. Given that an automorphism σ fixes the n_2-edges of $\tau(P)$, the two leaf edges of G_i cannot be interchanged, and thus are fixed and all edges of G_i are fixed. Thus every degree 2 vertex in $\tau'(P)$ is fixed as well. Finally, each degree 1 vertex in an edge $e \in H_i$ has a different label, and thus all these vertices are fixed.

Finally, since σ fixes each G_i componentwise, and the n_2-edge X_i intersects G_i more than any other n_2-edge, σ fixes each n_2-edge. □

Theorem 9.2. *Theorem 1.5 holds when* $m_2 = 2$.

Proof. All components of $\tau'(P)$ have a a nonpositive value by Lemma 4.7. By Lemma 4.7, if G is a tree of 0 value, then G has two leaves and is a chain. Furthermore, to assure asymmetry, the leaves of G must contain vertices labeled $\{1\}$ and $\{2\}$ respectively. Otherwise, if G is a non-tree with value 0, then since $r = 2$, G must have t edges and $2t$ vertices, and every vertex must be labeled [2].

We conclude that if $\tau'(P)$ has value $m_2 2^{m_2-1}$, then $\tau'(P)$ is a collection of chains, as described above; components in which every vertex is labeled [2]; and a degree 0 vertex of every nonempty label. But then $\tau(P)$ has a symmetry that results from reversing each chain and switching the n_2-edges. Thus the upper bound on n_2 holds.

Now we prove the sufficiency of the bound by construction. Let $\tau'(P)$ contain a chain with at least five edges, $e_1, \dots, e_{m_1}$ such that e_a and e_b intersect only when $b = a+1$, and then $e_a \cap e_b = \{v_a\}$. All vertices are labeled [2] except for v_1 and v_2, which are labeled $\{1\}$ and $\{2\}$ respectively, and degree 1 vertices u_1 and u_2 that are contained in each of the leaves, labeled $\{1\}$ and $\{2\}$ respectively. Also, $\tau'(P)$ contains a degree 0 vertex of each nonempty label.

We show that $\tau(P)$ is asymmetric. The n_2-edges cannot be switched since no automorphism of $\tau'(P)$ moves v_1 to any other vertex of weight 1. Thus v_1 is a fixed point in $\tau(P)$, which fixes all edges of $\tau'(P)$. Furthermore, the vertices in the leaves of $\tau'(P)$ are fixed since they are of different labels. □

Theorem 9.3. *Theorem 1.5 holds for odd* m_2.

Proof. By Lemma 4.7, no component of $\tau'(P)$ has positive value. We establish the upper bound on n_2 by showing that every connected component of $\tau'(P)$ has negative value.

Suppose that G is a component with value 0. By Lemma 4.4, G contains $|E(G)|(n_1-1)$ vertices, of which $|E(G)|(n_1-2)$ have degree 1. Furthermore, every edge contains n_1-2 degree 1 vertices since $w_{n_1-1} + w_{n_1-3} < 2w_{n_2-2}$. Construct G' by removing these degree 1 vertices. Then G' is a ordinary, 2-regular connected graph and thus a cycle. Furthermore, G lacks defects. We conclude that G has a nontrivial automorphism.

Now consider the following construction when $m_1 \geq m_2 + 3$. Let $\tau'(P)$ contain a connected n_1-uniform hypergraph with edges $e_1, \dots, e_{m_1}$, subscripts mod m_1, such that e_a and e_b do not intersect unless $|b-a| = 1$. If $b = a+1$, then we say that $e_a \cap e_b = \{v_a\}$. $\tau'(P)$ has no defects except for the following. For $1 \leq i \leq m_2 - 1$, say that v_i is labeled $[m_2] - \{i\}$, and v_{m_2+1} is labeled $[m_2] - \{m_2\}$. Also, $\tau'(P)$ contains a degree 0 vertex of each nonempty label. Then $\tau'(P)$ is asymmetric, and n_2 is the maximum value. □

10 $k \geq 1$ and $j = \lfloor (m_2 - 1)/2 \rfloor$

We say J that an (ψ, s, t, n_1)-regular hypergraph if J satisfies the following conditions. J is n_1-uniform with t edges, of which all edges intersect s other edges with the exception of ψ edges that each intersect $s - 1$ other edges. Furthermore, no two edges intersect at more than 1 vertex, and every vertex has degree at most 2. Finally, every automorphism of J fixes all edges. Before we prove the main result of this section, we need some lemmas on the existence of regular hypergraphs.

Lemma 10.1. *Let ψ and $s \geq 3$ be given, and suppose that t is sufficiently large relative to ψ and s. Suppose that $st - \psi$ is even. Then there exists an asymmetric graph with t vertices such that all vertices have degree s, except for ψ vertices that have degree $s - 1$.*

Call a graph of this form an (ψ, s, t)-*asymmetric* graph.

Proof. The lemma follows from the main theorem of [9] when st is even and $\psi = 0$, since an random s-regular graph is almost surely asymmetric. Such a graph is also almost surely s-connected [3].

Consider the case that st is even. Then ψ is also even. Choose distinct $t_1, \ldots, t_{\psi/2}$ such that $t_1 + \ldots + t_{\psi/2} = t$, with each t_i even if t is even. For $1 \leq i \leq \psi/2$, let G_i be an s-connected $(0, s, t_i)$-asymmetric graph with an edge removed. Then the disjoint union of the G_i is an (ψ, s, t)-asymmetric graph.

For odd st and ψ, we may construct a (ψ, s, t)-asymmetric graph as follows. Let G' be an $(\psi(s-1), s, t-\psi)$-asymmetric graph. Add new vertices $v_1, \ldots, v_\psi$ to G', each with disjoint neighbor sets of size $s - 1$ vertices of degree $s - 1$ in G'. The resulting graph is (ψ, s, t)-asymmetric. □

Lemma 10.2. *Let ψ and $s \geq 3$ be given, and suppose that t is sufficiently large relative to ψ and s. Suppose that $st - \psi$ is even. Also let $n_1 \geq s$ be given. Then there exists an (ψ, s, t, n_1)-asymmetric hypergraph.*

Proof. Let G be an (ψ, s, t)-asymmetric graph. Let H' be a hypergraph with vertex set $E(G)$, edge set $V(G)$, and incidence given by incident in G. Then construct H from H' by adding $n_1 - d$ degree 1 vertices to every edge in H' that contains d vertices. Then H is (ψ, s, t, n_1)-asymmetric. □

Proof of Theorem 1.6: The upper bound on n_2 follows by Corollary 4.6 and 4.8, except when km_1 is even and m_2 odd. In this case, suppose that all connected components of $\tau'(P)$ have value 0, and $\tau'(P)$ contains a degree 0 vertex of every nonempty label. By Lemma 4.4 and the fact that $w_{n_1-k-1} + w_{n_1-k-3} < 2w_{n_1-k-2}$, every edge of $\tau'(P)$ contains exactly $n_1 - k - 2$ degree 1 vertices. Furthermore, $\tau'(P)$ does not have any defects. In this case, $\tau(P)$ has an nontrivial automorphism that permutes the n_2-edges. The upper bound on n_2 follows.

We establish the result by the following constructions. First consider the case that km_1 and m_2 are both even. Let the graph of $\tau'(P)$ be an $(m_2, k+2, m_1, n_1)$-asymmetric hypergraph, which exists by Lemma 10.2, together with a degree 0 vertex of every nonempty label. Choose the labels of the vertices of $\tau'(P)$ so that there are no defects, label the edges with $k + 1$ degree 1 vertices by $e_1, \ldots, e_{m_2}$, and say e_i contains a vertex with label $\{i, i+1, \ldots, i + m_2/2 - 1\}$, subscripts mod m_2. Then $\tau'(P)$ is asymmetric and satisfies $n_2 = rm_1 + 2^{m_2-1}$.

If km_1 is even and m_2 is odd, let $\tau'(P)$ be a $(0, k+2, m_1, n_1)$-asymmetric hypergraph, together with a degree 0 vertex of every nonempty label. Suppose that $\tau'(P)$ has exactly the following m_2 defects: for degree 2 vertices $v_1, \ldots, v_{m_2}$, v_i is labeled $[m_2] - \{i\}$. Then $\tau(P)$ is asymmetric, and $n_2 = rm_1 + 2^{m_2-1} - 1$.

If km_1 is odd and m_2 is even, then let $\tau'(P)$ be an $(m_2+1, k+2, m_1, n_1)$-asymmetric hypergraph with $e_1, \ldots, e_{m_2+1}$ the edges that intersect $k+1$ other edges, together with a degree 0 vertex of every nonempty label. Choose the labels of the vertices of $\tau'(P)$ so that there are no defects, except that e_{m_2+1} contains a degree 1 vertex with every label of size at least $m_2 - j$, together with a vertex labeled $\emptyset$. For $1 \leq i \leq m_2$, e_i contains a degree 1 vertex labeled $\{i, i+1, \ldots, i+m_2/2-1\}$, subscripts mod m_2. Then $\tau(P)$ is asymmetric and satisfies $n_2 = rm_1 + 2^{m_2-1} - 1/2$.

Finally, if km_1 and m_2 are both odd, then let $\tau'(P)$ be an $(m_2, k+2, m_1, n_1)$-asymmetric hypergraph, together with a degree 0 vertex of every nonempty label. Choose the labels of $\tau'(P)$ so that there are no defects, and if the edges with $k+1$ degree 1 vertices are $e_1, \ldots, e_{m_2}$, then e_i contains a vertex labeled $\{i, i+1, \ldots, i+m_2/2-1/2\}$, with subscripts mod m_2. Then $\tau(P)$ is asymmetric and satisfies $n_2 = rm_1 + 2^{m_2-1} - 1/2$.

References

[1] M. Albertson, K. Collins, Symmetry breaking in graphs, *Electron. J. Combin.* **3** #R18 (1996), 17 pp.

[2] R. Bailey, P. Cameron, Base size, metric dimension and other invariants of groups and graphs, *Bulletin of the London Mathematical Society*, **43** (2011), 209–242.

[3] Béla Bollobás, *Random Graphs*, 2nd edition, Cambridge University Press (2001). Section 7.6.

[4] F. Bergeron, G. Labelle, P. Leroux, *Thorie des espces et combinatoire des structures arborescentes*, LaCIM, Montréal 1994. English version: *Combinatorial Species and Tree-like Structures*, Cambridge University Press, 1998.

[5] D. Boutin, Identifying graph automorphisms using determining sets, *Electron. J. Combin.*, **13(1)** #R78 (2006), 12 pp.

[6] D. Boutin, Small Label Classes in 2-Distinguishing Labelings, *Ars Math. Contemp.*, **1(2)** (2008), 154–164.

[7] M. Ellingham, J. Schroeder, Distinguishing partitions and asymmetric uniform hypergraphs, *Ars Math. Contemp.*, **4(1)** (2011), 111–123.

[8] F. Harary, R. W. Robinson, A. J. Schwenk, Twenty-step algorithm for determining the asymptotic number of trees of various species, *J. Austral. Math. Soc., Series A*, **20** (1975), 483–503. Errata: Vol. A **41** (1986), p. 325.

[9] J. H. Kim, B. Sudakov, V. Vu, *On the asymmetry of random regular graphs and random graphs*, Random Structures Algorithms - Special issue: *Proceedings of the tenth international conference "Random structures and algorithms"*, Volume 21 Issue 3-4, October 2002.

[10] R. Otter. The Number of Trees, *Ann. Math.* **49** (1948), 583-599.

[11] J. Tymoczko, Distinguishing numbers for graphs and groups, *Electron. J. Combin.* **11** #R63 (2004), 13 pp.

ARS MATHEMATICA CONTEMPORANEA

Also available at http://amc-journal.eu
ISSN 1855-3966 (printed edn.), ISSN 1855-3974 (electronic edn.)
ARS MATHEMATICA CONTEMPORANEA 10 (2016) 67–77

Commutators of cycles in permutation groups

Aleš Vavpetič *

Fakulteta za Matematiko in Fiziko, Univerza v Ljubljani,
Jadranska ulica 19, SI-1111 Ljubljana, Slovenija

Received 10 January 2013, accepted 26 September 2014, published online 27 May 2015

Abstract

We prove that for $n \geq 5$, every element of the alternating group A_n is a commutator of two cycles of A_n. Moreover we prove that for $n \geq 2$, a $(2n+1)$-cycle of the permutation group S_{2n+1} is a commutator of a p-cycle and a q-cycle of S_{2n+1} if and only if the following three conditions are satisfied (i) $n+1 \leq p, q$, (ii) $2n+1 \geq p, q$, (iii) $p+q \geq 3n+1$.

Keywords: Commutator, cycle, permutation, alternating group.

Math. Subj. Class.: 20B05

1 Introduction

In 1951 O. Ore [9] conjectured that in a finite simple non-abelian group every element is a commutator. In the same paper he proved that the conjecture holds for the alternating group A_n, where $n \geq 5$, but the result had already been proved by G. A. Miller half a century earlier [7]. After Ore published the paper there were many papers devoted to the Ore conjecture: R. C. Thompson proved the Ore conjecture for the projective special linear groups $PSL_n(q)$ [10], [11], [12], R. Gow proved it for the projective simplectic groups $PSp_{2n}(q)$, where $q \equiv 1 \pmod 4$ [4], O. Bonten for the exceptional groups of Lie type of low rank [2], J. Neubüser, H. Pahlings, E. Cleuvers proved it for the sporadic groups [8], E. W. Ellers, N. Gordeev handled the finite simple groups of Lie type over a finite field $\mathbb{F}_q$, whenever $q \geq 9$, ... M. W. Liebeck, E. A. O'Brien, A. Shalev, P. H. Tiep proved the Ore conjecture for the remaining cases [6] and the conjecture became the theorem. We refer the reader to the survey paper [5] for more historical notes about commutators and the Ore conjecture.

In this paper we prove a stronger version of the Ore conjecture for the simple alternating group A_n. In Section 2 it is shown that, for $n \geq 5$, every permutation of A_n is actually a

*This research was supported by the Slovenian Research Agency grants P1-0292-0101 and J1-5435-0101.
E-mail address: ales.vavpetic@fmf.uni-lj.si (Aleš Vavpetič)

commutator of two cycles of A_n. In particular, every even permutation of the symmetric group S_n is a product of two conjugate cycles. Namely, if $\rho = [\sigma, \tau] = \sigma^{-1}\tau^{-1}\sigma\tau$, then ρ is a product of $\sigma^{-1}\tau^{-1}\sigma$ and τ (and also a product of σ^{-1} and $\tau^{-1}\sigma\tau$). Note that permutations τ and τ^{-1} are conjugate in S_n. In [1] it is proved that a $(2n+1)$-cycle of A_{2n+1} is a product of two conjugate l-cycles of A_{2n+1} if and only if $l \geq n+1$. Hence this is a necessary condition for the existence of two l-cycles σ and τ such that $[\sigma, \tau]$ is a $(2n+1)$-cycle. In Section 3 it is shown that this is far from being a sufficient condition. More precisely, it is shown that, for $n \geq 2$, a $(2n+1)$-cycle of A_{2n+1} is a commutator of a p-cycle and a q-cycle of S_{2n+1} if and only if $n+1 \leq p, q$ and $p+q \geq 3n+1$. In particular, a $(2n+1)$-cycle of A_{2n+1} ($n \geq 2$) is a commutator of l-cycles of S_{2n+1} if and only if $l \geq \frac{3n+1}{2}$.

The image of an element a under a permutation σ is denoted by a^σ. Permutations are executed from left to right. The support $\operatorname{supp}\sigma$ of a permutation σ is the set of all elements which are not fixed by σ.

Let σ be a permutation, $a \in \operatorname{supp}\sigma$ and $x_1, \ldots, x_n \notin \operatorname{supp}\sigma$. We define permutations $\varphi(\sigma; a, x_1, \ldots, x_n)$ and $\varepsilon(\sigma; a)$ by

$$t^{\varphi(\sigma;a,x_1,\ldots,x_n)} = \begin{cases} x_1, & t = a, \\ x_{i+1}, & t \in \{x_1, \ldots, x_{n-1}\}, \\ a^\sigma, & t = x_n, \\ t^\sigma, & t \notin \{a, x_1, \ldots, x_n\}, \end{cases}$$

and

$$t^{\varepsilon(\sigma;a)} = \begin{cases} a, & t = a, \\ a^\sigma, & t = a^{\sigma^{-1}}, \\ t^\sigma, & t \notin \{a, a^{\sigma^{-1}}\}. \end{cases}$$

If σ is the k-cycle $(a_1, \ldots, a_k)$, then $\varphi(\sigma; a_k, x_1, \ldots, x_n)$ is the $(k+n)$-cycle $(a_1, \ldots, a_k, x_1, \ldots, x_n)$ and $\varepsilon(\sigma; a_k)$ is the $(k-1)$-cycle $(a_1, \ldots, a_{k-1})$.

Let σ and τ be permutations such that $\operatorname{supp}\sigma \cap \operatorname{supp}\tau = \emptyset$. For $a \in \operatorname{supp}\sigma$ and $b \in \operatorname{supp}\tau$, let $\psi(\sigma, \tau; a, b)$ denote the permutation defined by

$$t^{\psi(\sigma,\tau;a,b)} = \begin{cases} t^\sigma, & t \in \operatorname{supp}\sigma - \{a\}, \\ b^\tau, & t = a, \\ t^\tau, & t \in \operatorname{supp}\tau - \{b\}, \\ a^\sigma, & t = b. \end{cases}$$

If τ is a k-cycle then $\psi(\sigma, \tau; a, b) = \varphi(\sigma; a, b^\tau, b^{\tau^2}, \ldots, b^{\tau^k})$, and if σ is a k-cycle then $\psi(\sigma, \tau; a, b) = \varphi(\tau; b, a^\sigma, a^{\sigma^2}, \ldots, a^{\sigma^k})$.

2 Permutations as commutators of cycles

The proof that every permutation of A_n ($n \geq 5$) is a commutator of two cycles is based on induction on the number and the lengths of cycles in the cycle decomposition of the permutation. In the following lemmas we describe how the application of φ, ψ, and ε modify commutators.

Lemma 2.1. *Let σ, τ be permutations, $x \in \operatorname{supp}\sigma$, $y \in \operatorname{supp}\tau$, and $(\operatorname{supp}\sigma \cup \operatorname{supp}\tau) \cap (\{x_1, \dots, x_n\} \cup \{y_1, \dots, y_m\}) = \emptyset$. Then for $t \notin \{x^\sigma, x^{\tau\sigma}, y^{\tau\sigma}, y^{\sigma^{-1}\tau\sigma}, x_1, \dots, x_n, y_1, \dots, y_m\}$ we have $t^{[\sigma,\tau]} = t^{[\varphi(\sigma;x,x_1,\dots,x_n),\varphi(\tau;y,y_1,\dots,y_m)]}$.*

Proof. Denote $\widetilde{\sigma} = \varphi(\sigma; x, x_1, \dots, x_n)$ and $\widetilde{\tau} = \varphi(\tau; y, y_1, \dots, y_m)$. For $t \notin \{x^\sigma, x^{\tau\sigma}, y^{\tau\sigma}, y^{\sigma^{-1}\tau\sigma}, x_1, \dots, x_n, y_1, \dots, y_m\}$ we have $t^{\sigma^{-1}} = t^{\widetilde{\sigma}^{-1}}$. Since $t \notin \{y^{\tau\sigma}, y_1, \dots, y_m\}$, also $t^{\sigma^{-1}} \notin \{y^\tau, y_1, \dots, y_m\}$ and therefore $t^{\sigma^{-1}\tau^{-1}} = t^{\widetilde{\sigma}^{-1}\widetilde{\tau}^{-1}}$. Since $t^{\sigma^{-1}\tau^{-1}} \notin \{x, x_1, \dots, x_n\}$ we have $t^{\sigma^{-1}\tau^{-1}\sigma} = t^{\widetilde{\sigma}^{-1}\widetilde{\tau}^{-1}\widetilde{\sigma}}$. And finally $t^{\sigma^{-1}\tau^{-1}\sigma} \notin \{y, y_1, \dots, y_m\}$, hence $t^{[\sigma,\tau]} = t^{[\widetilde{\sigma},\widetilde{\tau}]}$. $\square$

We record the following immediate consequence.

Corollary 2.2. *Let σ, τ be permutations. Suppose that $a, b \in \operatorname{supp}\sigma$ such that $a^\sigma = a^\tau = b$, and $(\operatorname{supp}\sigma \cup \operatorname{supp}\tau) \cap (\{x_1, \dots, x_n\} \cup \{y_1, \dots, y_m\}) = \emptyset$. Then for $t \notin \{b^\sigma, b^{\tau\sigma}, x_1, \dots, x_n, y_1, \dots, y_m\}$ we have $t^{[\sigma,\tau]} = t^{[\varphi(\sigma;b,x_1,\dots,x_n),\varphi(\tau;b,y_1,\dots,y_m)]}$.*

Lemma 2.3. *Let σ, τ be permutations and $a, b \in \operatorname{supp}\sigma$ such that $b = a^\sigma = a^\tau$ and $c, d \notin \operatorname{supp}\sigma \cup \operatorname{supp}\tau$. Then*

$$[\varphi(\sigma; b, c, d), \varphi(\tau; b, d, c)] = \varphi([\sigma, \tau]; b^{\tau\sigma}, c, d).$$

Proof. Denote $\widetilde{\sigma} = \varphi(\sigma; b, c, d)$ and $\widetilde{\tau} = \varphi(\tau; b, d, c)$. By Corollary 2.2, we have $t^{[\widetilde{\sigma},\widetilde{\tau}]} = t^{[\sigma,\tau]}$ for $t \notin \{b^\sigma, b^{\tau\sigma}, c, d\}$. Because

$$\begin{aligned}
(b^{\tau\sigma})^{[\widetilde{\sigma},\widetilde{\tau}]} &= (b^\tau)^{\widetilde{\tau}^{-1}\widetilde{\sigma}\widetilde{\tau}} = c^{\widetilde{\sigma}\widetilde{\tau}} = d^{\widetilde{\tau}} = c,\\
c^{[\widetilde{\sigma},\widetilde{\tau}]} &= b^{\widetilde{\tau}^{-1}\widetilde{\sigma}\widetilde{\tau}} = a^{\widetilde{\sigma}\widetilde{\tau}} = b^{\widetilde{\tau}} = d,\\
d^{[\widetilde{\sigma},\widetilde{\tau}]} &= c^{\widetilde{\tau}^{-1}\widetilde{\sigma}\widetilde{\tau}} = d^{\widetilde{\sigma}\widetilde{\tau}} = (b^\sigma)^{\widetilde{\tau}} = b^{\sigma\tau} = (b^{\tau\sigma})^{[\sigma,\tau]},\\
(b^\sigma)^{[\widetilde{\sigma},\widetilde{\tau}]} &= d^{\widetilde{\tau}^{-1}\widetilde{\sigma}\widetilde{\tau}} = b^{\widetilde{\sigma}\widetilde{\tau}} = c^{\widetilde{\tau}} = b^\tau = (a^\sigma)^\tau = (b^{\tau^{-1}})^{\sigma\tau} = (b^\sigma)^{[\sigma,\tau]},
\end{aligned}$$

we have $[\varphi(\sigma; b, c, d), \varphi(\tau; b, d, c)] = \varphi([\sigma, \tau]; b^{\tau\sigma}, c, d)$. $\square$

Lemma 2.4. *Let σ, τ be permutations and $a, b \in \operatorname{supp}\sigma$ such that $b = a^\sigma = a^\tau$ and $c, d \notin \operatorname{supp}\sigma \cup \operatorname{supp}\tau$. Then*

$$\begin{aligned}
[\varphi(\sigma; b, c, d), \varphi(\tau; b, d)] &= \varphi([\sigma, \tau]; b^\sigma, c, d),\\
[\varphi(\sigma; b, d), \varphi(\tau; b, c, d)] &= \varphi([\sigma, \tau]; b^\sigma, d, c).
\end{aligned}$$

Proof. Denote $\widetilde{\sigma} = \varphi(\sigma; b, c, d)$ and $\widetilde{\tau} = \varphi(\tau; b, d)$. By Corollary 2.2, we have $t^{[\widetilde{\sigma},\widetilde{\tau}]} = t^{[\sigma,\tau]}$ for $t \notin \{b^\sigma, b^{\tau\sigma}, c, d\}$. Because

$$\begin{aligned}
(b^\sigma)^{[\widetilde{\sigma},\widetilde{\tau}]} &= d^{\widetilde{\tau}^{-1}\widetilde{\sigma}\widetilde{\tau}} = b^{\widetilde{\sigma}\widetilde{\tau}} = c^{\widetilde{\tau}} = c,\\
c^{[\widetilde{\sigma},\widetilde{\tau}]} &= b^{\widetilde{\tau}^{-1}\widetilde{\sigma}\widetilde{\tau}} = a^{\widetilde{\sigma}\widetilde{\tau}} = b^{\widetilde{\tau}} = d,\\
d^{[\widetilde{\sigma},\widetilde{\tau}]} &= c^{\widetilde{\tau}^{-1}\widetilde{\sigma}\widetilde{\tau}} = c^{\widetilde{\sigma}\widetilde{\tau}} = d^{\widetilde{\tau}} = b^\tau = (a^\sigma)^\tau = (b^{\tau^{-1}})^{\sigma\tau} = (b^\sigma)^{[\sigma,\tau]},\\
(b^{\tau\sigma})^{[\widetilde{\sigma},\widetilde{\tau}]} &= (b^\tau)^{\widetilde{\tau}^{-1}\widetilde{\sigma}\widetilde{\tau}} = d^{\widetilde{\sigma}\widetilde{\tau}} = (b^\sigma)^{\widetilde{\tau}} = b^{\sigma\tau} = (b^{\tau\sigma})^{[\sigma,\tau]},
\end{aligned}$$

we have $[\varphi(\sigma; b, c, d), \varphi(\tau; b, d)] = \varphi([\sigma, \tau]; b^\sigma, c, d)$.

Because $[\sigma,\tau]^{-1}=[\tau,\sigma]$ and $(b^\tau)^{[\tau,\sigma]}=b^\sigma$, we have

$$\begin{aligned}[\varphi(\sigma;b,d),\varphi(\tau;b,c,d)] &= ([\varphi(\tau;b,c,d),\varphi(\sigma;b,d)])^{-1} = \\ &= \varphi([\tau,\sigma];b^\tau,c,d)^{-1} = \\ &= \varphi([\sigma,\tau];b^\sigma,d,c).\end{aligned}$$

□

Corollary 2.5. *Let ρ be a $(2n+1)$-cycle and $n\geq 2$. For $p,q\in\mathbb{N}$ such that $p,q\leq 2n+1$ and $p+q\geq 3n+2$, there exist a p-cycle σ, a q-cycle τ, and $a\in\operatorname{supp}\sigma$ such that $[\sigma,\tau]=\rho$, $\operatorname{supp}\rho=\operatorname{supp}\sigma\cup\operatorname{supp}\tau$, and $a^\sigma=a^\tau$. In the case $q\neq 2n+1$ we arrange that $a^{\sigma\sigma}\notin\operatorname{supp}\tau$.*

Proof. If $n=2$ and $p\geq q$ then $(p,q)\in\{(5,5),(5,4),(5,3),(4,4)\}$ and we have

$$\begin{aligned}(a_1,a_2,a_3,a_4,a_5) &= [(a_1,a_4,a_2,a_3,a_5),(a_1,a_4,a_3,a_5,a_2)] = \\ &= [(a_1,a_4,a_2,a_5,a_3),(a_1,a_4,a_3,a_5)] = \\ &= [(a_1,a_2,a_4,a_5,a_3),(a_1,a_2,a_5)] = \\ &= [(a_1,a_5,a_2,a_3),(a_1,a_5,a_3,a_4)].\end{aligned}$$

If $n=2$ and $p<q$, then $q=2n+1=5$ and we can use the equality $[\sigma,\tau]^{-1}=[\tau,\sigma]$. In all cases $a_1^\sigma=a_1^\tau$ and if $q\neq 5$, also $a_1^{\sigma\sigma}\notin\operatorname{supp}\tau$.

Let $n>2$. The proof is divided into 3 cases.

Case 1: Suppose $q\leq 2n$. Let $p_1=p-2$, $q_1=q-1$, and $n_1=n-1$. Then $p_1+q_1=p-2+q-1\geq 3n_1+2$ and $p_1,q_1\leq 2n_1+1$. By the inductive hypothesis there exist a p_1-cycle σ, a q_1-cycle τ, and $a\in\operatorname{supp}\sigma$ such that $[\sigma,\tau]$ is a $(2n_1+1)$-cycle, $\operatorname{supp}\sigma\cup\operatorname{supp}\tau=\operatorname{supp}[\sigma,\tau]$, and $a^\sigma=a^\tau$. Let $x,y\notin\operatorname{supp}\sigma\cup\operatorname{supp}\tau$, $\widetilde{\sigma}=\varphi(\sigma;a^\sigma,x,y)$, and $\widetilde{\tau}=\varphi(\tau;a^\tau,y)$. Then $\widetilde{\sigma}$ is a p-cycle, $\widetilde{\tau}$ is a q-cycle, $a^{\widetilde{\sigma}}=a^\sigma=a^\tau=a^{\widetilde{\tau}}$, $a^{\widetilde{\sigma}\widetilde{\sigma}}=x\notin\operatorname{supp}\widetilde{\tau}$, and by Lemma 2.4, $[\widetilde{\sigma},\widetilde{\tau}]$ is a $(2n+1)$-cycle and $\operatorname{supp}\widetilde{\sigma}\cup\operatorname{supp}\widetilde{\tau}=\operatorname{supp}[\widetilde{\sigma},\widetilde{\tau}]$.

Case 2: Suppose $q=2n+1$ and $p\neq 2n+1$. This case follows from the previous case and equality $[\sigma,\tau]^{-1}=[\tau,\sigma]$.

Case 3: Suppose $p=q=2n+1$. By the inductive hypothesis there exist $(2n-1)$-cycles σ, τ, and $a\in\operatorname{supp}\sigma$ such that $[\sigma,\tau]$ is a $(2n-1)$-cycle, $\operatorname{supp}\sigma=\operatorname{supp}\tau=\operatorname{supp}[\sigma,\tau]$, and $a^\sigma=a^\tau$. Let $x,y\notin\operatorname{supp}\sigma$, $\widetilde{\sigma}=\varphi(\sigma;a^\sigma,x,y)$, and $\widetilde{\tau}=\varphi(\tau;a^\tau,y,x)$. Then $\widetilde{\sigma}$ and $\widetilde{\tau}$ are $(2n+1)$-cycles, $a^{\widetilde{\sigma}}=a^\sigma=a^\tau=a^{\widetilde{\tau}}$, and by Lemma 2.3, $[\widetilde{\sigma},\widetilde{\tau}]$ is a $(2n+1)$-cycle and $\operatorname{supp}\widetilde{\sigma}=\operatorname{supp}\widetilde{\tau}=\operatorname{supp}[\widetilde{\sigma},\widetilde{\tau}]$. □

Lemma 2.6. *Let σ,τ be permutations and $a,b\in\operatorname{supp}\sigma$ such that $b=a^\sigma=a^\tau$, $b^\sigma\notin\operatorname{supp}\tau$, and $c\notin\operatorname{supp}\sigma\cup\operatorname{supp}\tau$. Then*

$$[\sigma,\varphi(\tau;b,c)]=\varepsilon([\sigma,\tau];b^\sigma)(c,b^\sigma).$$

Proof. Let $\widetilde{\tau}=\varphi(\tau;b,c)$. By Corollary 2.2, we get $t^{[\sigma,\widetilde{\tau}]}=t^{[\sigma,\tau]}$ for $t\notin\{b^\sigma,b^{\tau\sigma},c\}$. From

$$(b^\sigma)^{[\sigma,\widetilde{\tau}]}=b^{\widetilde{\tau}^{-1}\sigma\widetilde{\tau}}=a^{\sigma\widetilde{\tau}}=b^{\widetilde{\tau}}=c,$$

$$c^{[\sigma,\widetilde{\tau}]}=c^{\widetilde{\tau}^{-1}\sigma\widetilde{\tau}}=b^{\sigma\widetilde{\tau}}=b^\sigma,$$

$$(b^{\tau\sigma})^{[\sigma,\widetilde{\tau}]}=(b^\tau)^{\widetilde{\tau}^{-1}\sigma\widetilde{\tau}}=c^{\sigma\widetilde{\tau}}=c^{\widetilde{\tau}}=b^\tau,$$

and

$$(b^{\tau\sigma})^{[\sigma,\tau]} = b^{\sigma\tau} = b^{\sigma},$$
$$(b^{\sigma})^{[\sigma,\tau]} = b^{\tau^{-1}\sigma\tau} = a^{\sigma\tau} = b^{\tau},$$

it follows $[\sigma, \varphi(\tau; b, c)] = \varepsilon([\sigma,\tau]; b^{\sigma})(c, b^{\sigma})$. □

Corollary 2.7. *Let $n_1, n_2 \in \mathbb{N}$ and let ρ be a product of two disjoint cycles of lengths $2n_1$ and $2n_2$, respectively. If $p, q \leq 2(n_1+n_2)-1$ and $p+q \geq 3(n_1+n_2)$ then there exist a p-cycle σ, a q-cycle τ, and $a \in \operatorname{supp}\sigma$ such that $\rho = [\sigma,\tau]$, $\operatorname{supp}\rho = \operatorname{supp}\sigma \cup \operatorname{supp}\tau$, and $a^{\sigma} = a^{\tau}$.*

If $n_1 = n_2 = 1$ then there exist no cycles σ and τ such that the length of one of them is strictly greater than $2(n_1+n_2)-1 = 3$, $[\sigma,\tau]$ is a product of two disjoint transpositions, $\operatorname{supp}[\sigma,\tau] = \operatorname{supp}\sigma \cup \operatorname{supp}\tau$, where $a^{\sigma} = a^{\tau}$ for some $a \in \operatorname{supp}\sigma$. That means that in the Corollary in this case the upper bound requirement on the length of the cycles is sharp. If $n_1+n_2 \geq 3$ the upper bound requirement is not sharp (it can be increased to $2(n_1+n_2)$) but the bound in the Corollary is in almost all cases sufficient for our purposes. Namely, in the case $n_1+n_2 \geq 4$, we get $2(2(n_1+n_2)-2) \geq 3(n_1+n_2)$ and therefore the Corollary provides two cycles whose lengths can be required to be (independently) either odd or even: both odd ($p = q = 2(n_1+n_2)-1$), both even ($p = q = 2(n_1+n_2)-2$), the first even and the second odd ($p = 2(n_1+n_2)-2, q = 2(n_1+n_2)-1$), the first odd and the second even.

Proof. One may assume that $n_1 \geq n_2$. The proof is by induction on n_2.

Let $n_2 = 1$. If $n_1 = 1$ then the only possibility for p and q is $p = q = 3$. In this case $[(a_1,a_2,a_3),(a_1,a_2,a_4)] = (a_1,a_2),(a_3,a_4)$. Let $n_1 \geq 2$. Because $p+(q-1) \geq 3(n_1+1)-1 = 3n_1+2$ and $p,q \leq 2(n_1+1)-1 = 2n_1+1$, Corollary 2.5 provides a p-cycle σ, a $(q-1)$-cycle τ, and $a \in \operatorname{supp}\sigma$ such that $[\sigma,\tau]$ is a $(2n_1+1)$-cycle, $\operatorname{supp}\sigma \cup \operatorname{supp}\tau = \operatorname{supp}[\sigma,\tau]$, $a^{\sigma} = a^{\tau}$, and $a^{\sigma\sigma} \notin \operatorname{supp}\tau$. Let $c \notin \operatorname{supp}\sigma \cup \operatorname{supp}\tau$ and $\widetilde{\tau} = \varphi(\tau; a^{\tau}, c)$. Then $\widetilde{\tau}$ is a q-cycle, $a^{\sigma} = a^{\tau} = a^{\widetilde{\tau}}$, and by Lemma 2.6, $[\sigma,\widetilde{\tau}] = \varepsilon([\sigma,\tau]; a^{\sigma\sigma})(a^{\sigma\sigma}, c)$ and $\operatorname{supp}\sigma \cup \operatorname{supp}\widetilde{\tau} = \operatorname{supp}[\sigma,\widetilde{\tau}]$. Note that $a^{\sigma\widetilde{\tau}\sigma} = c$ is in the support of the 2-cycle.

For the proof by induction, suppose that for all $n < n_2$ the assumptions $p,q \leq 2(n_1+n)-1$ and $p+q \geq 3(n_1+n)$ guarantee the existence of a p-cycle σ, a q-cycle τ, and $a \in \operatorname{supp}\sigma$ such that the following hold: $[\sigma,\tau]$ is a product of two disjoint cycles of lengths $2n_1$ and $2n$, $\operatorname{supp}[\sigma,\tau] = \operatorname{supp}\sigma \cup \operatorname{supp}\tau$, $a^{\sigma} = a^{\tau}$, and $a^{\sigma\tau\sigma}$ is in the support of the $2m$-cycle in the cycle decomposition of $[\sigma,\tau]$.

We prove that the same holds for $n = n_2$. The proof is divided into 3 cases.

Case 1: Let $q < 2(n_1+n_2)-1$. Define $\widetilde{p} = p-2$, $\widetilde{q} = q-1$, and $m = n_2-1$. Because $\widetilde{p}+\widetilde{q} \geq 3(n_1+m)$ and $\widetilde{p},\widetilde{q} \leq 2(n_1+m)-1$, the inductive hypothesis yields a $\widetilde{p}$-cycle σ, a $\widetilde{q}$-cycle τ, and $a \in \operatorname{supp}\sigma$ such that $[\sigma,\tau] = \rho_1\rho_2$, where $\operatorname{supp}\rho_1 \cap \operatorname{supp}\rho_2 = \emptyset$, ρ_1 is a $2n_1$-cycle, ρ_2 is a $2m$-cycle, $a^{\sigma} = a^{\tau}$, and $a^{\sigma\tau\sigma} \in \operatorname{supp}\rho_2$. Let $x, y \notin \operatorname{supp}\sigma \cup \operatorname{supp}\tau$, $\widetilde{\sigma} = \varphi(\sigma; a^{\sigma}, x, y)$, and $\widetilde{\tau} = \varphi(\tau; a^{\tau}, y)$. Then $\widetilde{\sigma}$ is a p-cycle, $\widetilde{\tau}$ is a q-cycle, $a^{\widetilde{\sigma}} = a^{\sigma} = a^{\tau} = a^{\widetilde{\tau}}$, and by Lemma 2.4, $[\widetilde{\sigma},\widetilde{\tau}] = \varphi(\rho_1\rho_2; a^{\sigma\sigma}, x, y) = \rho_1\varphi(\rho_2; a^{\sigma\sigma}, x, y)$ and $a^{\widetilde{\sigma}\widetilde{\tau}\widetilde{\sigma}} = a^{\sigma\sigma} \in \operatorname{supp}\varphi(\rho_2; a^{\sigma\sigma}, x, y)$.

Case 2: Let $p \neq 2(n_1+n_2)-1$ and $q = 2(n_1+n_2)-1$. This case follows from the previous case and the equality $[\sigma,\tau]^{-1} = [\tau,\sigma]$.

Case 3: Let $p = q = 2(n_1+n_2)-1$. Define $\widetilde{p} = \widetilde{q} = 2(n_1+n_2)-3$ and $m = n_2 - 1$. From $\widetilde{p}, \widetilde{q} \le 2(n_1+m)-1$ and $n_1 > 1$ we get $\widetilde{p}+\widetilde{q} \ge 3(n_1+m)$. By the inductive hypothesis there exist $\widetilde{p}$-cycles σ, τ, and $a \in \operatorname{supp}\sigma$ such that $[\sigma,\tau] = \rho_1\rho_2$, where $\operatorname{supp}\rho_1 \cap \operatorname{supp}\rho_2 = \emptyset$, ρ_1 is a $2n_1$-cycle, ρ_2 is a $2m$-cycle, $a^\sigma = a^\tau$, and $a^{\sigma\tau\sigma} \in \operatorname{supp}\rho_2$. Let $x, y \notin \operatorname{supp}\sigma \cup \operatorname{supp}\tau$, $\widetilde{\sigma} = \varphi(\sigma; a^\sigma, x, y)$, and $\widetilde{\tau} = \varphi(\tau; a^\tau, y, x)$. Then $\widetilde{\sigma}$ and $\widetilde{\tau}$ are p-cycles, $a^{\widetilde{\sigma}} = a^\sigma = a^\tau = a^{\widetilde{\tau}}$, and by Lemma 2.3, $[\widetilde{\sigma},\widetilde{\tau}] = \varphi(\rho_1\rho_2; a^{\sigma\tau\sigma}, x, y) = \rho_1\varphi(\rho_2; a^{\sigma\tau\sigma}, x, y)$ and $a^{\widetilde{\sigma}\widetilde{\tau}\widetilde{\sigma}} = a^{\sigma\sigma} \in \operatorname{supp}\varphi(\rho_2; a^{\sigma\tau\sigma}, x, y)$. □

Lemma 2.8. *Let σ, τ be permutations and $a, b \in \operatorname{supp}\sigma$ such that $b = a^\sigma = a^\tau$, and $x, y, z \notin \operatorname{supp}\sigma \cup \operatorname{supp}\tau$. Then*

$$[\varphi(\sigma; b, x, y, z), \varphi(\tau; b, y, z)] = [\sigma,\tau](x, y, z).$$

Proof. Let $\widetilde{\sigma} = \varphi(\sigma; b, x, y, z)$ and $\widetilde{\tau} = \varphi(\tau; b, y, z)$. By Corollary 2.2, we have $t^{[\widetilde{\sigma},\widetilde{\tau}]} = t^{[\sigma,\tau]}$ for $t \notin \{b^\sigma, b^{\tau\sigma}, x, y, z\}$. As

$$\begin{aligned}
(b^\sigma)^{[\widetilde{\sigma},\widetilde{\tau}]} &= z^{\widetilde{\tau}^{-1}\widetilde{\sigma}\widetilde{\tau}} = y^{\widetilde{\sigma}\widetilde{\tau}} = z^{\widetilde{\tau}} = b^\tau = (a^\sigma)^\tau = (b^{\tau^{-1}})^{\sigma\tau} = (b^\sigma)^{[\sigma,\tau]},\\
(b^{\tau\sigma})^{[\widetilde{\sigma},\widetilde{\tau}]} &= (b^\tau)^{\widetilde{\tau}^{-1}\widetilde{\sigma}\widetilde{\tau}} = z^{\widetilde{\sigma}\widetilde{\tau}} = b^{\sigma\widetilde{\tau}} = b^{\sigma\tau} = (b^{\tau\sigma})^{[\sigma,\tau]},\\
x^{[\widetilde{\sigma},\widetilde{\tau}]} &= b^{\widetilde{\tau}^{-1}\widetilde{\sigma}\widetilde{\tau}} = a^{\widetilde{\sigma}\widetilde{\tau}} = b^{\widetilde{\tau}} = y,\\
y^{[\widetilde{\sigma},\widetilde{\tau}]} &= x^{\widetilde{\tau}^{-1}\widetilde{\sigma}\widetilde{\tau}} = x^{\widetilde{\sigma}\widetilde{\tau}} = y^{\widetilde{\tau}} = z,\\
z^{[\widetilde{\sigma},\widetilde{\tau}]} &= y^{\widetilde{\tau}^{-1}\widetilde{\sigma}\widetilde{\tau}} = b^{\widetilde{\sigma}\widetilde{\tau}} = x^{\widetilde{\tau}} = x,
\end{aligned}$$

we have $[\widetilde{\sigma},\widetilde{\tau}] = [\sigma,\tau](x, y, z)$. □

Lemma 2.9. *Let $\sigma_1, \sigma_2, \tau_1, \tau_2$ be cycles such that $(\operatorname{supp}\sigma_1 \cup \operatorname{supp}\tau_1) \cap (\operatorname{supp}\sigma_2 \cup \operatorname{supp}\tau_2) = \emptyset$. Suppose there exist $a \in \operatorname{supp}\sigma_1$ and $b \in \operatorname{supp}\sigma_2$ such that $a^{\sigma_1} = a^{\tau_1}$ and $b^{\sigma_2} = b^{\tau_2}$. Then $[\psi(\sigma_1, \sigma_2; a^{\sigma_1}, b^{\sigma_2}), \psi(\tau_1, \tau_2; a^{\tau_1}, b^{\tau_2})] = [\sigma_1, \tau_1][\sigma_2, \tau_2]$.*

Proof. Let $\sigma = \psi(\sigma_1, \sigma_2; a^{\sigma_1}, b^{\sigma_2})$ and $\tau = \psi(\tau_1, \tau_2; a^{\tau_1}, b^{\tau_2})$. Set $c = a^{\sigma_1} = a^{\tau_1}$ and $d = b^{\sigma_2} = b^{\tau_2}$. From Corollary 2.2 and equalities $\sigma = \varphi(\sigma_1; c, b^{\sigma_2^2}, \dots, b, b^{\sigma_2})$ and $\tau = \varphi(\tau_1; c, b^{\tau_2^2}, \dots, b, b^{\tau_2})$, we get $t^{[\sigma,\tau]} = t^{[\sigma_1,\tau_1]} = t^{[\sigma_1,\tau_1][\sigma_2,\tau_2]}$ for $t \notin \{c^{\sigma_1}, c^{\tau_1\sigma_1}\} \cup \operatorname{supp}\sigma_2 \cup \operatorname{supp}\tau_2$. From Corollary 2.2 and equalities $\sigma = \varphi(\sigma_2; d, a^{\sigma_1^2}, \dots, a, a^{\sigma_1})$ and $\tau = \varphi(\tau_2; d, a^{\tau_1^2}, \dots, a, a^{\tau_1})$, we get $t^{[\sigma,\tau]} = t^{[\sigma_2,\tau_2]} = t^{[\sigma_1,\tau_1][\sigma_2,\tau_2]}$ for $t \notin \{d^{\sigma_2}, d^{\tau_2\sigma_1}\} \cup \operatorname{supp}\sigma_2 \cup \operatorname{supp}\tau_2$. Therefore $t^{[\sigma,\tau]} = t^{[\sigma_1,\tau_1][\sigma_2,\tau_2]}$ for $t \notin \{c^{\sigma_1}, c^{\tau_1\sigma_1}, d^{\sigma_2}, d^{\tau_2\sigma_1}\}$. From

$$\begin{aligned}
(c^{\sigma_1})^{[\sigma,\tau]} &= d^{\tau^{-1}\sigma\tau} = b^{\sigma\tau} = d^\tau = c^{\tau_1} = a^{\sigma_1\tau_1} = c^{\tau_1^{-1}\sigma_1\tau_1} = (c^{\sigma_1})^{[\sigma_1,\tau_1]},\\
(c^{\tau_1\sigma_1})^{[\sigma,\tau]} &= (c^{\tau_1})^{\tau^{-1}\sigma\tau} = d^{\sigma\tau} = (c^{\sigma_1})^\tau = c^{\sigma_1\tau_1} = (c^{\tau_1\sigma_1})^{[\sigma_1,\tau_1]},\\
(d^{\sigma_2})^{[\sigma,\tau]} &= c^{\tau^{-1}\sigma\tau} = a^{\sigma\tau} = c^\tau = d^{\tau_2} = b^{\sigma_2\tau_2} = d^{\tau_2^{-1}\sigma_2\tau_2} = (d^{\sigma_2})^{[\sigma_2,\tau_2]},\\
(d^{\tau_2\sigma_2})^{[\sigma,\tau]} &= (d^{\tau_2})^{\tau^{-2}\sigma\tau} = c^{\sigma\tau} = (d^{\sigma_2})^\tau = d^{\sigma_2\tau_2} = (d^{\tau_2\sigma_2})^{[\sigma_2,\tau_2]},
\end{aligned}$$

we get $[\sigma,\tau] = [\sigma_1,\tau_1][\sigma_2,\tau_2]$. □

Theorem 2.10. *Let $\rho \in A_n$. If $n \ge 5$ or ρ is not a 3-cycle then ρ is a commutator of two cycles of A_n.*

Proof. If $\rho = (a_1, a_2, a_3)$ is a 3-cycle then $n \geq 5$ and $\rho = [(a_1, a_3, x), (a_1, a_2, y)]$ for some $x, y \notin \operatorname{supp} \rho$.

Suppose that ρ is not a 3-cycle. We show that there exist cycles σ and τ of odd lengths and $a \in \operatorname{supp} \sigma$ such that $\rho = [\sigma, \tau]$, $\operatorname{supp} \rho = \operatorname{supp} \sigma \cup \operatorname{supp} \tau$, and $a^\sigma = a^\tau$. The proof is by induction on the number of cycles in the cycle decomposition of ρ, which we denote by $c(\rho)$.

If $c(\rho) = 1$, ρ is a cycle of odd length $l \geq 5$. The statement follows from Corollary 2.5.

If $c(\rho) = 2$, then let $\rho = \rho_1\rho_2$, where ρ_1 and ρ_2 are disjoint cycles. The lengths of these cycles are of the same parity. If the lengths are even, the statement follows from Corollary 2.7. In the case of odd lengths, 3 cases are considered.

Case 1: Suppose both lengths are 3. Then $[(a_1, a_2, a_6, a_5, a_3), (a_1, a_2, a_4, a_6, a_5)] = (a_1, a_2, a_3)(a_4, a_5, a_6)$.

Case 2: Suppose exactly one of the lengths is 3. One may assume $\rho_2 = (x, y, z)$ is the 3-cycle. Let ρ_1 be a cycle of length $2l + 1$, where $l \geq 2$. By Corollary 2.5, there exist a $2l$-cycle σ, a $(2l + 1)$-cycle τ, and $a \in \operatorname{supp} \sigma$ such that $\rho_1 = [\sigma, \tau]$, $\operatorname{supp} \rho_1 = \operatorname{supp} \sigma \cup \operatorname{supp} \tau$, and $a^\sigma = a^\tau$. By Lemma 2.8, we have $\rho = [\varphi(\sigma; a^\sigma, x, y, z), \varphi(\tau; a^\tau, y, z)]$, where $\varphi(\sigma; a^\sigma, x, y, z)$ and $\varphi(\tau; a^\tau, y, z)$ are $(2l + 3)$-cycles.

Case 3: Suppose both lengths are greater than 3. Let ρ_i be a cycle of length $2l_i + 1$, $l_i \geq 2$. By Corollary 2.5, there exist $(2l_1 + 1)$-cycles σ_1, τ_1, $(2l_2)$-cycles σ_2, τ_2, $a_1 \in \operatorname{supp} \sigma_1$, and $a_2 \in \operatorname{supp} \sigma_2$ such that $\rho_i = [\sigma_i, \tau_i]$, $\operatorname{supp} \rho_i = \operatorname{supp} \sigma_i \cup \operatorname{supp} \tau_i$, and $a_i^{\sigma_i} = a_i^{\tau_i}$. Then $\psi(\sigma_1, \sigma_2; a_1^{\sigma_1}, a_2^{\sigma_2})$ and $\psi(\tau_1, \tau_2; a_1^{\tau_1}, a_2^{\tau_2})$ are $(2(l_1 + l_2) + 1)$-cycles and by Lemma 2.9, $\rho = [\psi(\sigma_1, \sigma_2; a_1^{\sigma_1}, a_2^{\sigma_2}), \psi(\tau_1, \tau_2; a_1^{\tau_1}, a_2^{\tau_2})]$.

If $c(\rho) \geq 3$, the following 4 cases are considered.

Case 1: Suppose $\rho = \rho_1\rho_2$, where ρ_2 is a $(2l+1)$-cycle, $l \geq 2$, and $\operatorname{supp} \rho_1 \cap \operatorname{supp} \rho_2 = \emptyset$. By Corollary 2.5, there exist $(2l)$-cycles σ_2, τ_2 and $b \in \operatorname{supp} \sigma_2$, such that $\rho_2 = [\sigma_2, \tau_2]$, $\operatorname{supp} \rho_2 = \operatorname{supp} \sigma_2 \cup \operatorname{supp} \tau_2$, and $b^{\sigma_2} = b^{\tau_2}$. Because $2 \leq c(\rho_1) \leq c(\rho) - 1$, the inductive hypothesis yields cycles σ_1, τ_1 of odd lengths, as well as $a \in \operatorname{supp} \sigma_1$, such that $\rho_1 = [\sigma_1, \tau_1]$, $\operatorname{supp} \rho_1 = \operatorname{supp} \sigma_1 \cup \operatorname{supp} \tau_1$, and $a^{\sigma_1} = a^{\tau_1}$. By Lemma 2.9, we have $\rho = [\psi(\sigma_1, \sigma_2; a^{\sigma_1}, b^{\sigma_2}), \psi(\tau_1, \tau_2; a^{\tau_1}, b^{\tau_2})]$, where $\psi(\sigma_1, \sigma_2; a^{\sigma_1}, b^{\sigma_2})$ and $\psi(\tau_1, \tau_2; a^{\tau_1}, b^{\tau_2})$ are cycles of odd lengths.

Case 2: Suppose $\rho = \rho_1\rho_2$, where $\rho_2 = (a_1, a_2, a_3)(a_4, a_5, a_6)$ and $\operatorname{supp} \rho_1 \cap \operatorname{supp} \rho_2 = \emptyset$. If $\rho_1 = (a_7, a_8, a_9)$ then $\rho = [(a_1, a_2, a_7, a_8, a_9, a_4, a_5, a_3, a_6), (a_1, a_2, a_8, a_9, a_5, a_3, a_4)]$. If ρ_1 is not a 3-cycle, the inductive hypothesis yields cycles σ_1, τ_1 of odd lengths, as well as $a \in \operatorname{supp} \sigma_1$, such that $\rho_1 = [\sigma_1, \tau_1]$, $\operatorname{supp} \rho_1 = \operatorname{supp} \sigma_1 \cup \operatorname{supp} \tau_1$, and $a^{\sigma_1} = a^{\tau_1} = b$. Then $\sigma = \varphi(\varphi(\sigma_1; b, a_1, a_2, a_3); b, a_4, a_5, a_6)$ and $\tau = \varphi(\varphi(\tau_1; b, a_2, a_3); b, a_5, a_6)$ are cycles of odd lengths and, using Lemma 2.8 twice, we get $\rho = [\sigma, \tau]$.

Case 3: Suppose $\rho = \rho_1\rho_2$, where ρ_2 is a disjoint product of cycles of lengths $2l_1$ and $2l_2$, such that $l_1 + l_2 \geq 3$, and $\operatorname{supp} \rho_1 \cap \operatorname{supp} \rho_2 = \emptyset$.

If $\rho_1 = (a_1, a_2, a_3)$ then by Corollary 2.7, there exist a $(2(l_1 + l_2) - 2)$-cycle σ_2, a $(2(l_1 + l_2) - 1)$-cycle τ_2, and $a \in \operatorname{supp} \sigma_2$, such that $\rho_2 = [\sigma_2, \tau_2]$, $\operatorname{supp} \rho_2 = \operatorname{supp} \sigma_2 \cup \operatorname{supp} \tau_2$, and $a^{\sigma_2} = a^{\tau_2} = b$. Then $\sigma = \varphi(\sigma_2; b, a_1, a_2, a_3)$ and $\tau = \varphi(\tau_2; b, a_2, a_3)$ are $(2(l_1 + l_2) + 1)$-cycles and by Lemma 2.8, we get $\rho = [\sigma, \tau]$.

If ρ_1 is not a 3-cycle then by the inductive hypothesis there exist cycles σ_1, τ_1 of odd lengths and $a \in \operatorname{supp} \sigma_1$, such that $\rho_1 = [\sigma_1, \tau_1]$, $\operatorname{supp} \rho_1 = \operatorname{supp} \sigma_1 \cup \operatorname{supp} \tau_1$, and $a^{\sigma_1} = a^{\tau_1}$. If $l_1 + l_2 = 3$ then $\rho_2 = (a_1, a_2, a_3, a_4)(a_5, a_6)$ and for $\sigma_2 = (a_1, a_5, a_2, a_4, a_6, a_3)$ and $\tau_2 = (a_1, a_5, a_3, a_4)$ we get $\rho_2 = [\sigma_2, \tau_2]$ and for $b = a_1$ we get $b^{\sigma_1} = b^{\tau_1}$. If $l_1 + l_2 > 3$ Corollary 2.7 provides $(2(l_1 + l_2) - 2)$-cycles σ_2 and τ_2, as well as $b \in \operatorname{supp} \sigma_2$, such that

$\rho_2 = [\sigma_2, \tau_2]$, $\operatorname{supp}\rho_2 = \operatorname{supp}\sigma_2 \cup \operatorname{supp}\tau_2$, and $b^{\sigma_2} = b^{\tau_2}$. Then $\sigma = \psi(\sigma_1, \sigma_2; a^{\sigma_1}, b^{\sigma_2})$ and $\tau = \psi(\tau_1, \tau_2; a^{\tau_1}, b^{\tau_2})$ are cycles of odd length and by Lemma 2.9, we get $\rho = [\sigma, \tau]$.

Case 4: Suppose ρ is a disjoint product of transpositions and at most one 3-cycle. If there are at most four transpositions in the cycle decomposition of ρ we have 3 possibilities:

$$
\begin{aligned}
&[(a_1,a_3,a_5,a_6,a_2,a_4,a_7),(a_1,a_3,a_6,a_4,a_7,a_2,a_5)] = (a_1,a_2)(a_3,a_4)(a_5,a_6,a_7),\\
&[(a_1,a_2,a_4,a_8,a_6,a_3,a_5),(a_1,a_2,a_3,a_8,a_4,a_6,a_7)] = (a_1,a_2)(a_3,a_4)(a_5,a_6)(a_7,a_8),\\
&[(a_1,a_2,a_5,a_3,a_4,a_9,a_{10},a_7,a_{11}),(a_1,a_2,a_6,a_3,a_4,a_{10},a_7,a_9,a_8)] =\\
&\quad =(a_1,a_2)(a_3,a_4)(a_5,a_6)(a_7,a_8)(a_9,a_{10},a_{11}).
\end{aligned}
$$

Otherwise $\rho = \rho_1\rho_2$, where $\rho_2 = (a_1, a_2)(a_3, a_4)(a_5, a_6)(a_7, a_8)$, $2 \le c(\rho_1) < c(\rho)$, and $\operatorname{supp}\rho_1 \cap \operatorname{supp}\rho_2 = \emptyset$. By the inductive hypothesis there exist cycles σ_1, τ_1 of odd lengths and $a \in \operatorname{supp}\sigma_1$, such that $\rho_1 = [\sigma_1, \tau_1]$, $\operatorname{supp}\rho_1 = \operatorname{supp}\sigma_1 \cup \operatorname{supp}\tau_1$, and $a^{\sigma_1} = a^{\tau_1}$. For $\sigma_2 = (a_1, a_8, a_3, a_2, a_4, a_6, a_7, a_5)$ and $\tau_2 = (a_1, a_8, a_4, a_3, a_5, a_6)$ we have $\rho_2 = [\sigma_2, \tau_2]$. Then $\sigma = \psi(\sigma_1, \sigma_2; a^{\sigma_1}, a_1^{\sigma_2})$ and $\tau = \psi(\tau_1, \tau_2; a^{\tau_1}, a_1^{\tau_2})$ are cycles of odd lengths and by Lemma 2.9, we get $\rho = [\sigma, \tau]$. □

3 Cycles as commutators of cycles

From the previous section we know that a $(2n + 1)$-cycle is a commutator of a p-cycle and a q-cycle if $p + q \ge 3n + 2$ (and $p, q \le 2n + 1$). But this sufficient condition is not necessary. Note that in the previous section we were interested in pairs of cycles σ and τ, for which there exists $a \in \operatorname{supp}\sigma$ such that $a^\sigma = a^\tau$. We needed that for "concatenation" of cycles in Lemma 2.9. With that assumption withdrawn, the result is obtained by using a more stringent hypothesis as shown in the next corollary.

Lemma 3.1. *Let σ, τ be permutations, $x, y \notin \operatorname{supp}\sigma \cup \operatorname{supp}\tau$, $a_1, a_2 \in \operatorname{supp}\sigma \cap \operatorname{supp}\tau$, $b \in \operatorname{supp}\sigma - \operatorname{supp}\tau$, and $c \in \operatorname{supp}\tau - \operatorname{supp}\sigma$, such that $a_1^\sigma = b$, $b^\sigma = a_2$, $a_1^\tau = c$, and $c^\tau = a_2$. Then*

$$[\varphi(\sigma; b, c, x), \varphi(\tau; c, y)] = \varphi([\sigma, \tau]; c, y, x).$$

Proof. Let $\widetilde{\sigma} = \varphi(\sigma; b, c, x)$ and $\widetilde{\tau} = \varphi(\tau; c, y)$. If $t \notin \{x, a_2, c\}$ then $t^{\sigma^{-1}} = t^{\widetilde{\sigma}^{-1}}$. If $t \notin \{y, a_2^\sigma\}$ then $t^{\sigma^{-1}} \notin \{y, a_2\}$ and $t^{\sigma^{-1}\tau^{-1}} = t^{\sigma^{-1}\widetilde{\tau}^{-1}}$. If $t \notin \{x, a_2^\sigma, a_2\}$ then $t^{\sigma^{-1}\tau^{-1}} \notin \{x, c, b\}$ and $t^{\sigma^{-1}\tau^{-1}\sigma} = t^{\sigma^{-1}\tau^{-1}\widetilde{\sigma}}$. If $t \notin \{y, a_2^\sigma\}$ then $t^{\sigma^{-1}\tau^{-1}\sigma} \notin \{y, c\}$ and $t^{\sigma^{-1}\tau^{-1}\sigma\tau} = t^{\sigma^{-1}\tau^{-1}\widetilde{\sigma}\widetilde{\tau}}$. Hence for $t \notin \{x, y, c, a_2, a_2^\sigma\}$ we get $t^{[\sigma,\tau]} = t^{[\widetilde{\sigma},\widetilde{\tau}]}$. Because

$$
\begin{aligned}
c^{[\sigma,\tau]} &= c^{\tau^{-1}\sigma\tau} = a_1^{\sigma\tau} = b^\tau = b,\\
c^{[\widetilde{\sigma},\widetilde{\tau}]} &= b^{\widetilde{\tau}^{-1}\widetilde{\sigma}\widetilde{\tau}} = b^{\widetilde{\sigma}\widetilde{\tau}} = c^{\widetilde{\tau}} = y,\\
y^{[\widetilde{\sigma},\widetilde{\tau}]} &= y^{\widetilde{\tau}^{-1}\widetilde{\sigma}\widetilde{\tau}} = c^{\widetilde{\sigma}\widetilde{\tau}} = x^{\widetilde{\tau}} = x,\\
x^{[\widetilde{\sigma},\widetilde{\tau}]} &= c^{\widetilde{\tau}^{-1}\widetilde{\sigma}\widetilde{\tau}} = a_1^{\widetilde{\sigma}\widetilde{\tau}} = b^{\widetilde{\tau}} = b,\\
a_2^{[\widetilde{\sigma},\widetilde{\tau}]} &= x^{\widetilde{\tau}^{-1}\widetilde{\sigma}\widetilde{\tau}} = x^{\widetilde{\sigma}\widetilde{\tau}} = a_2^{\widetilde{\tau}} = a_2^{\tau} = b^{\sigma\tau} = b^{\tau^{-1}\sigma\tau} = a_2^{[\sigma,\tau]},\\
(a_2^\sigma)^{[\widetilde{\sigma},\widetilde{\tau}]} &= a_2^{\widetilde{\tau}^{-1}\widetilde{\sigma}\widetilde{\tau}} = y^{\widetilde{\sigma}\widetilde{\tau}} = y^{\widetilde{\tau}} = a_2 = c^\tau = c^{\sigma\tau} = a_2^{\tau^{-1}\sigma\tau} = (a_2^\sigma)^{[\sigma,\tau]},
\end{aligned}
$$

we get $[\widetilde{\sigma}, \widetilde{\tau}] = \varphi([\sigma, \tau]; c, y, x)$. □

Corollary 3.2. *Let ρ be a $(2n+1)$-cycle and $n \geq 2$. For $p, q \in \mathbb{N}$ such that $p, q \leq 2n$ and $p + q = 3n + 1$, there exist a p-cycle σ and a q-cycle τ, such that $[\sigma, \tau] = \rho$ and* $\operatorname{supp}\rho = \operatorname{supp}\sigma \cup \operatorname{supp}\tau$.

Proof. By induction on n we prove that whenever $p, q \leq 2n$ and $p + q = 3n + 1$, there exist a p-cycle σ, a q-cycle τ, $a_1, a_2 \in \operatorname{supp}\sigma \cap \operatorname{supp}\tau$, $b \in \operatorname{supp}\sigma - \operatorname{supp}\tau$, and $c \in \operatorname{supp}\tau - \operatorname{supp}\sigma$, such that $a_1^\sigma = b$, $b^\sigma = a_2$, $a_1^\tau = c$, $c^\tau = a_2$, $[\sigma, \tau]$ is a $(2n+1)$-cycle, and $\operatorname{supp}[\sigma, \tau] = \operatorname{supp}\sigma \cup \operatorname{supp}\tau$.

Because $[\tau, \sigma] = [\sigma, \tau]^{-1}$ we may assume $p \geq q$.

If $n = 2$ then $p = 4$, $q = 3$ and we have $[(a_1, b, a_2, d), (a_1, c, a_2)] = (a_1, c, b, d, a_2)$.

Let $n > 2$. For $p, q \leq 2n$ and $p + q = 3n + 1$ we define $\widetilde{p} = p - 2$ and $\widetilde{q} = q - 1$. Then $\widetilde{p} + \widetilde{q} = 3(n-1) + 1$ and $\widetilde{p} \leq 2(n-1)$. From $q \leq p$ we get $q \neq 2n$ and therefore $\widetilde{q} \leq 2(n-1)$. By the inductive hypothesis there exist a $\widetilde{p}$-cycle $\widetilde{\sigma}$, a $\widetilde{q}$-cycle $\widetilde{\tau}$, $a_1, a_2 \in \operatorname{supp}\widetilde{\sigma} \cap \operatorname{supp}\widetilde{\tau}$, $b \in \operatorname{supp}\widetilde{\sigma} - \operatorname{supp}\widetilde{\tau}$, and $c \in \operatorname{supp}\widetilde{\tau} - \operatorname{supp}\widetilde{\sigma}$, such that $a_1^{\widetilde{\sigma}} = b$, $b^{\widetilde{\sigma}} = a_2$, $a_1^{\widetilde{\tau}} = c$, $c^{\widetilde{\tau}} = a_2$, $[\widetilde{\sigma}, \widetilde{\tau}]$ is a $(2n-1)$-cycle, and $\operatorname{supp}[\widetilde{\sigma}, \widetilde{\tau}] = \operatorname{supp}\widetilde{\sigma} \cup \operatorname{supp}\widetilde{\tau}$. Let $x, y \notin \operatorname{supp}\widetilde{\sigma} \cup \operatorname{supp}\widetilde{\tau}$. Then $\sigma = \varphi(\widetilde{\sigma}; b, c, x)$ is a p-cycle, $\tau = \varphi(\widetilde{\tau}; c, y)$ is a q-cycle, $c, a_2 \in \operatorname{supp}\sigma \cap \operatorname{supp}\tau$, $x \in \operatorname{supp}\sigma - \operatorname{supp}\tau$, $y \in \operatorname{supp}\tau - \operatorname{supp}\sigma$, $c^\sigma = x$, $x^\sigma = a_2$, $c^\tau = y$, $y^\tau = a_2$, and by Lemma 3.1, $[\sigma, \tau]$ is a $(2n+1)$-cycle. □

Let σ and τ be permutations. An equivalence relation on the set $\operatorname{supp}\sigma \cap \operatorname{supp}\tau$ is defined in the following way. Elements $a, b \in \operatorname{supp}\sigma \cap \operatorname{supp}\tau$ are equivalent if and only if there exist $a_0, \ldots, a_n \in \operatorname{supp}\sigma \cap \operatorname{supp}\tau$ and $\rho_1, \ldots, \rho_n \in \{\sigma, \sigma^{-1}, \tau, \tau^{-1}\}$, such that $a = a_0$, $b = a_n$, and $a_i = a_{i-1}^{\rho_i}$ for $i = 1, \ldots, n$. This is obviously an equivalence relation.

Definition 3.3. Permutations σ and τ are **braided** if all elements of $\operatorname{supp}\sigma \cap \operatorname{supp}\tau$ are equivalent to each other.

Lemma 3.4. *Let σ and τ be cycles such that the commutator $[\sigma, \tau]$ is a cycle and* $\operatorname{supp}[\sigma, \tau] = \operatorname{supp}\sigma \cup \operatorname{supp}\tau$. *Then σ and τ are braided.*

Proof. Let $\rho = [\sigma, \tau]$ and $a_0 \in \operatorname{supp}\sigma \cap \operatorname{supp}\tau$. For $n \geq 0$ we inductively define $a_{4n+1} = a_{4n}^{\sigma^{-1}}$, $a_{4n+2} = a_{4n+1}^{\tau^{-1}}$, $a_{4n+3} = a_{4n+2}^{\sigma}$, and $a_{4n+4} = a_{4n+3}^{\tau}$. Let us show that if $a_{4m} = a_0^{\rho^m} \in \operatorname{supp}\sigma \cap \operatorname{supp}\tau$, then a_{4m} is equivalent to a_0. Let $b_1 = a_0$ and $i_1 = \max\{i \mid i \leq 4m, a_i = a_0\}$. For $k \geq 1$ and $i_k < 4m$ we let $i_{k+1} = \max\{i \mid i_k < i \leq 4m, a_i = a_{i_k+1}\}$, $b_{k+1} = a_{i_{k+1}}$, and $\rho_k \in \{\sigma, \sigma^{-1}, \tau, \tau^{-1}\}$, where ρ_k is uniquely defined by $b_k^{\rho_k} = b_{k+1}$. If we show that $b_k \in \operatorname{supp}\sigma \cap \operatorname{supp}\tau$ for all k, then by definition, $a_0 = b_1$ is equivalent to $a_{4m} = b_l$. For $1 \leq k < l$ we have $b_{k+1} \in \operatorname{supp}\rho_k$. Suppose $b_{k+1} \notin \operatorname{supp}\widetilde{\rho}$, where $\widetilde{\rho}$ is the cycle in $\{\sigma, \tau\} - \{\rho_k, \rho_k^{-1}\}$. Because $a_{i_k}^{\rho_k} = a_{i_k+1}$ and $\rho_k \neq \widetilde{\rho}^{\pm 1}$, necessarily also $a_{i_k+1}^{\widetilde{\sigma}} = a_{i_k+2}$ or $a_{i_k+1}^{\widetilde{\sigma}^{-1}} = a_{i_k+2}$. Because $a_{i_k+2}^{\rho_k^{-1}} = a_{i_k+3}$ and $a_{i_k+1} \notin \operatorname{supp}\widetilde{\sigma}$, we get $a_{i_k} = a_{i_k+3}$. This contradicts the definition of i_k. Hence $b_k \in \operatorname{supp}\sigma \cap \operatorname{supp}\tau$.

Let $b \in \operatorname{supp}\sigma \cap \operatorname{supp}\tau$. Because ρ is a cycle and $b \in \operatorname{supp}\rho$, there exists m such that $b = a_0^{\rho^m}$. Thus b is equivalent to a_0, and hence σ and τ are braided. □

Lemma 3.5. *Let σ and τ be permutations such that* $\operatorname{supp}[\sigma, \tau] = \operatorname{supp}\sigma \cup \operatorname{supp}\tau$. *Then* $|\operatorname{supp}\sigma - \operatorname{supp}\tau|, |\operatorname{supp}\tau - \operatorname{supp}\sigma| \leq |\operatorname{supp}\sigma \cap \operatorname{supp}\tau|$.

Proof. Suppose there exist $x, y \in \operatorname{supp}\sigma - \operatorname{supp}\tau$, such that $x = y^\sigma$. Then $x^{[\sigma,\tau]} = x$, and consequently $x \notin \operatorname{supp}[\sigma, \tau]$, which is a contradiction. Hence the map $(\operatorname{supp}\sigma -$

$\operatorname{supp}\tau) \to (\operatorname{supp}\sigma \cap \operatorname{supp}\tau)$, defined by $x \mapsto x^\sigma$, is an injection. Therefore $|\operatorname{supp}\sigma - \operatorname{supp}\tau| \le |\operatorname{supp}\sigma \cap \operatorname{supp}\tau|$.

Because $\operatorname{supp}[\tau,\sigma] = \operatorname{supp}[\sigma,\tau]$, the other inequality follows from the above paragraph. □

Lemma 3.6. *Let σ and τ be cycles such that $[\sigma,\tau]$ is a cycle and $\operatorname{supp}[\sigma,\tau] = \operatorname{supp}\sigma \cup \operatorname{supp}\tau$. Then $|\operatorname{supp}\sigma - \operatorname{supp}\tau| + |\operatorname{supp}\tau - \operatorname{supp}\sigma| \le |\operatorname{supp}\sigma \cap \operatorname{supp}\tau| + 1$.*

Proof. Let $k = |\operatorname{supp}\sigma \cap \operatorname{supp}\tau|$, $|\operatorname{supp}\sigma| = k+p$, and $|\operatorname{supp}\tau| = k+q$. If $p = 0$, then by Lemma 3.5 we have

$$|\operatorname{supp}\sigma - \operatorname{supp}\tau| + |\operatorname{supp}\tau - \operatorname{supp}\sigma| = |\operatorname{supp}\tau - \operatorname{supp}\sigma| < |\operatorname{supp}\sigma \cap \operatorname{supp}\tau| + 1.$$

Analogously for $q = 0$. Let $p, q > 0$. Let $\operatorname{supp}\sigma - \operatorname{supp}\tau = \{a_1,\dots,a_p\}$. Let $m_i \in \mathbb{N} \cup \{0\}$ be the largest number such that $a_i^{\sigma^j} \in \operatorname{supp}\sigma \cap \operatorname{supp}\tau$ for all $j \in \{1,\dots,m_i\}$. We claim that all m_i are positive. Indeed, suppose that there exist $x, y \in \operatorname{supp}\sigma - \operatorname{supp}\tau$, such that $x^\sigma = y$. Then $y^{[\sigma,\tau]} = y$ which is a contradiction since $\operatorname{supp}\sigma \subset \operatorname{supp}[\sigma,\tau]$. Hence the set $M_i = \{a_i^\sigma,\dots,a_i^{\sigma^{m_i}}\}$ is nonempty for all i. Because σ is a cycle and $p > 0$, for every $x \in \operatorname{supp}\sigma \cap \operatorname{supp}\tau$ there exists the smallest $i \in \mathbb{N}$ such that $x^{\sigma^{-i}} = a_k$ for some k, which means that $x \in M_k$. Therefore, $(\operatorname{supp}\sigma \cap \operatorname{supp}\tau) = M_1 \coprod \dots \coprod M_p$. Similarly, $(\operatorname{supp}\sigma \cap \operatorname{supp}\tau) = N_1 \coprod \dots \coprod N_q$, where $\operatorname{supp}\tau - \operatorname{supp}\sigma = \{b_1,\dots,b_q\}$, $N_i = \{b_i^\tau,\dots,b_i^{\tau^{n_i}}\} \subset \operatorname{supp}\tau \cap \operatorname{supp}\sigma$, and $b_i^{\tau^{n_i+1}} \notin \operatorname{supp}\sigma$.

By Lemma 3.4, the cycles σ and τ are braided. Hence there exist $i_2 \in \{2,\dots,p\}$, $d_2 \in M_1$, $c_2 \in M_{i_2}$, and $\tau_2 \in \{\tau,\tau^{-1}\}$ such that $d_2 = c_2^{\tau_2}$. For $j > 2$ there exist $i_j \in \{2,\dots,p\} - \{i_2,\dots,i_{j-1}\}$, $d_j \in M_1 \cup (\cup_{l=2}^{j-1} M_{i_l})$, $c_j \in M_{i_j}$, and $\tau_j \in \{\tau,\tau^{-1}\}$ such that $d_j = c_j^{\tau_j}$. Let us show that for each i, the set $\widetilde{N}_i = N_i - \{c_2,\dots,c_p\}$ is nonempty. By construction, the elements $c_2,\dots,c_p$ are different, $d_j \ne c_k$ for $j \le k$, and every pair $\{c_j, d_j\}$ is a subset of N_l for some l. Suppose $N_i \cap \{c_2,\dots,c_p\} = \{c_{k_1},\dots,c_{k_r}\}$, where $k_1 < \dots < k_r$. Then $d_{k_1} \in N_i$ and $d_{k_1} \notin \{c_{k_1},\dots,c_{k_r}\}$, so $d_{k_1} \in \widetilde{N}_i \ne \emptyset$. Hence in the union of the q nonempty sets $\widetilde{N}_1,\dots,\widetilde{N}_q$ there are exactly $k-(p-1)$ elements. This means that $|\operatorname{supp}\tau - \operatorname{supp}\sigma| = q \le k-(p-1) = |\operatorname{supp}\sigma \cap \operatorname{supp}\tau| - |\operatorname{supp}\sigma - \operatorname{supp}\tau| + 1$. □

Theorem 3.7. *Let $n \ge 2$ and let ρ be a $(2n+1)$-cycle. There exist a p-cycle σ and a q-cycle τ such that $\rho = [\sigma,\tau]$ and $\operatorname{supp}\rho = \operatorname{supp}\sigma \cup \operatorname{supp}\tau$ if and only if the following three conditions are satisfied (i) $n+1 \le p, q$, (ii) $2n+1 \ge p, q$, (iii) $p+q \ge 3n+1$.*

Proof. Suppose there exist a p-cycle σ and a q-cycle τ such that $\rho = [\sigma,\tau]$ and $\operatorname{supp}\rho = \operatorname{supp}\sigma \cup \operatorname{supp}\tau$. Let $k = |\operatorname{supp}\sigma \cap \operatorname{supp}\tau|$, $p = k + \widetilde{p}$, and $q = k + \widetilde{q}$. By Lemma 3.5, we have $\widetilde{q} \le k$, therefore $2\widetilde{q} \le k + \widetilde{q} = q \le 2n+1$ which implies $\widetilde{q} \le n$. Then $2n+1 = |\operatorname{supp}\rho| = |\operatorname{supp}\sigma \cup \operatorname{supp}\tau| = p + \widetilde{q} \le p + n$, hence $n+1 \le p$. By Lemma 3.6, we have $\widetilde{p} + \widetilde{q} \le k+1$. Therefore $2n+1 = k + \widetilde{p} + \widetilde{q} \le 2k+1$ and $p + q = 2n+1+k \ge 3n+1$.

If $p+q \ge 3n+2$ the theorem follows from Corollary 2.5. If $p+q = 3n+1$, the theorem follows from Corollary 3.2. □

References

[1] E. Bertram, Even permutations as a product of two conjugate cycles, *J. Combin. Theory* **12** (1972), 368–380.

[2] O. Bonten, *Über Kommutatoren in endlichen einfachen Gruppen*, Aachener Beiträge zur Math. **7**, Verlag der Augustinus-Buchhandlung, Aachen, 1993.

[3] E. W. Ellers, N. Gordeev, On the conjectures of J. Thompson and O. Ore, *Trans. Amer. Math. Soc.* **350** (1998), no. 9, 3657–3671.

[4] R. Gow, Commutators in the symplectic group, *Arch. Math. (Basel)* **50** (1988), no. 3, 204–209.

[5] L. C. Kappe, R. S. Morse, *On commutators in groups*, Groups St. Andrews 2005. Vol. 2, 531–558, *London Math. Soc. Lecture Note Ser.*, 340, Cambridge Univ. Press, Cambridge, 2007.

[6] M. W. Liebeck, E. A. O'Brien, A. Shalev, P. H. Tiep, The Ore conjecture, *J. Eur. Math. Soc. (JEMS)* **12** (2010), no. 4, 939–1008.

[7] G. A. Miller, On the commutators of a given group, *Bull. Amer. Math. Soc.* **6** (1899) 105–109.

[8] J. Neubüser, H. Pahlings, E. Cleuvers, Each sporadic finasig G has a class C such that $CC = G$, *Abstracts Amer. Math. Soc.* **34**, 6 (1984).

[9] O. Ore, Some remarks on commutators, *Proc. Amer. Math. Soc.* **272** (1951), 307–314.

[10] R. C. Thompson, Commutators in the special and general linear groups, *Trans. Amer. Math. Soc.* **101** (1961), 16–33.

[11] R. C. Thompson, Commutators of matrices with coefficients from the field of two elements, *Duke Math. J.* **29** (1962), 367–373.

[12] R. C. Thompson, On matrix commutators, *Portugal. Math.* **21** (1962), 143–153.

ARS MATHEMATICA CONTEMPORANEA

Also available at http://amc-journal.eu
ISSN 1855-3966 (printed edn.), ISSN 1855-3974 (electronic edn.)
ARS MATHEMATICA CONTEMPORANEA 10 (2016) 79–84

An improvement of a result of Zverovich–Zverovich

Grant Cairns [*], Stacey Mendan
Department of Mathematics and Statistics, La Trobe University, Melbourne, Australia 3086

Received 8 March 2013, accepted 20 October 2013, published online 27 May 2015

Abstract

We give an improvement of a result of Zverovich and Zverovich which gives a condition on the first and last elements in a decreasing sequence of positive integers for the sequence to be graphic, that is, the degree sequence of a finite graph.

Keywords: Graph, graphic sequence.

Math. Subj. Class.: 05C07

1 Statement of Results

A finite sequence of positive integers is *graphic* if it occurs as the sequence of vertex degrees of a graph. Here, graphs are understood to be *simple*, in that they have no loops or repeated edges. A result of Zverovich and Zverovich states:

Theorem 1.1 ([8, Theorem 6]). *Let a, b be reals. If $\underline{d} = (d_1, \dots, d_n)$ is a sequence of positive integers in decreasing order with $d_1 \leq a, d_n \geq b$ and*

$$n \geq \frac{(1+a+b)^2}{4b},$$

then $\underline{d}$ is graphic.

Notice that here the term $\frac{(1+a+b)^2}{4b}$ is monotonic increasing in a, for $a \geq 1$ and fixed b, and it is also monotonic decreasing in b, for $a \geq b \geq 1$ and fixed a. Thus any sequence that satisfies the inequality $n \geq \frac{(1+a+b)^2}{4b}$, for any pair $a \geq d_1, b \leq d_n$, will also satisfy the inequality $n \geq \frac{(1+d_1+d_n)^2}{4d_n}$. So Theorem 1.1 has the following equivalent expression.

[*]Corresponding author
E-mail addresses: G.Cairns@latrobe.edu.au (Grant Cairns), spmendan@students.latrobe.edu.au (Stacey Mendan)

Theorem 1.2. *Suppose that $\underline{d} = (d_1, \ldots, d_n)$ is a decreasing sequence of positive integers with even sum. If*

$$n \geq \frac{(1 + d_1 + d_n)^2}{4d_n}, \tag{1.1}$$

then $\underline{d}$ is graphic.

The simplified form of Theorem 1.2 also affords a somewhat simpler proof, which we give in Section 2 below. Admittedly, the proof in [8] is already quite elementary, though it does use the strong index results of [4, 3].

The following corollary of Zverovich–Zverovich's is obtained by taking $a = d_1$ and $b = 1$ in Theorem 1.1.

Corollary 1.3 ([8, Corollary 2]). *Suppose that $\underline{d} = (d_1, \ldots, d_n)$ is a decreasing sequence of positive integers with even sum. If $d_1 \leq 2n^{\frac{1}{2}} - 2$, then $\underline{d}$ is graphic.*

Note that this can be expressed in the following equivalent form.

Corollary 1.4. *Suppose that $\underline{d} = (d_1, \ldots, d_n)$ is a decreasing sequence of positive integers with even sum. If $n \geq \frac{d_1^2}{4} + d_1 + 1$, then $\underline{d}$ is graphic.*

Zverovich–Zverovich state that the bound of Corollary 1.4 "cannot be improved", and they give examples to this effect. In fact, there is an improvement, as we will now describe. The subtlety here is that Zverovich–Zverovich formulated their result as an upper bound on d_1, and, as an upper bound on d_1, this upper bound on d_1 cannot be improved. However, the reformulation of their result as a lower bound on n can be slightly improved. We prove the following result in Section 2.

Theorem 1.5. *Suppose that $\underline{d} = (d_1, \ldots, d_n)$ is a decreasing sequence of positive integers with even sum. If $n \geq \left\lfloor \frac{d_1^2}{4} + d_1 \right\rfloor$, then $\underline{d}$ is graphic.*

Example 1.6. There are many examples of sequences that verify the hypotheses of Theorem 1.5 but not those of Corollary 1.4. In fact, there are 81 such sequences of length $n \leq 8$. Figure 1 shows three graphs whose degree sequences have this property; they have degree sequences $(2, 2, 2), (3, 3, 2, 2, 2)$ and $(3, 3, 3, 3, 3, 3)$ respectively. For infinite families of examples, for every positive odd integer x, consider the sequence $(2x, 1^{x^2+2x-1})$, and for x even, consider the sequence $(2x, 2x, 1^{x^2+2x-2})$. Here, and in sequences throughout this paper, the superscripts indicate the number of repetitions of the entry.

Example 1.7. The following examples show that the bound of Theorem 1.5 is sharp when $d_n = 1$. For d even, say $d = 2x$ with $x \geq 1$, let $\underline{d} = (d^{x+1}, 1^{x^2+x-2})$. For d odd, say $d = 2x + 1$ with $x \geq 1$, let $\underline{d} = (d^{x+1}, 1^{x^2+2x-1})$. In each case g has even sum, $n = \left\lfloor \frac{d^2}{4} + d \right\rfloor - 1$, but $\underline{d}$ is not graphic, as one can see from the Erdős–Gallai Theorem [6].

Remark 1.8. The fact that Theorem 1.2 is not sharp has also been remarked in [1], in the abstract of which the authors state that Theorem 1.2 is "sharp within 1". They give the bound

$$n \geq \frac{(1 + d_1 + d_n)^2 - \epsilon'}{4d_n}, \tag{1.2}$$

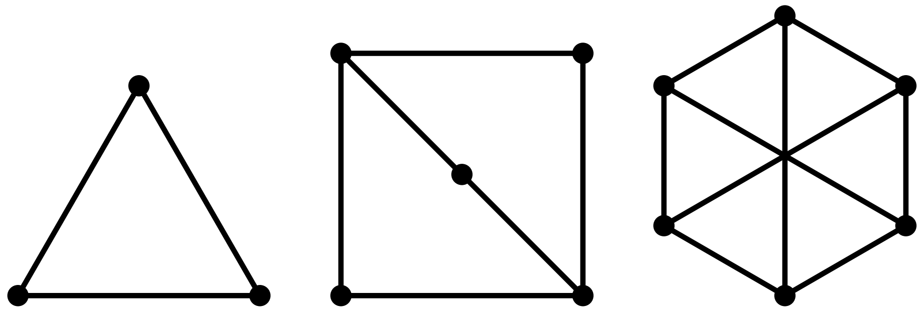


Figure 1: Three examples

where $\epsilon' = 0$ if $d_1 + d_n$ is odd, and $\epsilon' = 1$ otherwise. Consider any decreasing sequence with $d_1 = 2x+1$ and $d_n = 1$. Note that the bound given by Theorem 1.2 is $n \geq x^2+3x+3$, the bound given by (1.2) is $n \geq x^2 + 3x + 2$, while Theorem 1.5 gives the stronger bound $n \geq x^2 + 3x + 1$. The paper [1] gives more precise bounds, as a function of d_1, d_n, and the maximal gap in the sequence.

Remark 1.9. There are many other recent papers on graphic sequences; see for example [5, 7, 1, 2].

2 Proofs of Theorems 1.2 and 1.5

We will require the Erdős–Gallai Theorem, which we recall for convenience.

Erdős–Gallai Theorem. A sequence $\underline{d} = (d_1, \dots, d_n)$ of nonnegative integers in decreasing order is graphic if and only if its sum is even and, for each integer k with $1 \leq k \leq n$,

$$\sum_{i=1}^{k} d_i \leq k(k-1) + \sum_{i=k+1}^{n} \min\{k, d_i\}. \tag{EG}$$

Proof of Theorem 1.2. Suppose that $\underline{d} = (d_1, \dots, d_n)$ is a decreasing sequence with even sum, satisfying (1.1), and which is not graphic. By the Erdős–Gallai Theorem, there exists k with $1 \leq k \leq n$, such that

$$\sum_{i=1}^{k} d_i > k(k-1) + \sum_{i=k+1}^{n} \min\{k, d_i\}. \tag{2.1}$$

For each i with $1 \leq i \leq k$, replace d_i by d_1; the left hand side of (2.1) is not decreased, while the right hand side of (2.1) is unchanged, so (2.1) still holds. Now for each i with $k+1 \leq i \leq n$, replace d_i by d_n; the left hand side of (2.1) is unchanged, while the right hand side of (2.1) has not increased, so (2.1) again holds. Notice that if $k < d_n$, then (2.1) gives $kd_1 > k(k-1) + (n-k)k = k(n-1)$, and so $d_1 \geq n$. Then (1.1) would give $4nd_n \geq (1+d_n+n)^2$, that is, $(n-(d_n-1))^2-(d_n-1)^2+(1+d_n)^2 \leq 0$. But this inequality clearly has no solutions. Hence $k \geq d_n$. Thus (2.1) now reads $kd_1 > k(k-1)+(n-k)d_n$, or equivalently

$$(k - \frac{1}{2}(1 + d_1 + d_n))^2 - \frac{1}{4}(1 + d_1 + d_n)^2 + nd_n < 0.$$

But this contradicts the hypothesis. □

The following proof uses the same general strategy as the preceding proof, but requires a somewhat more careful argument.

Proof of Theorem 1.5. Suppose that $\underline{d}$ satisfies the hypotheses of the theorem. First suppose that d_1 is even, say $d_1 = 2x$. If $d_n \geq 2$, then since $\frac{(1+d_n+d_1)^2}{4d_n}$ is a strictly monotonic decreasing function of d_n for $1 \leq d_n \leq d_1$, we have

$$n \geq \frac{d_1^2}{4} + d_1 = \frac{(2+d_1)^2}{4} - 1 > \frac{(1+d_n+d_1)^2}{4d_n} - 1,$$

so $n \geq \frac{(1+d_n+d_1)^2}{4d_n}$ and hence $\underline{d}$ is graphic by Theorem 1.2. So, assuming that $\underline{d}$ is not graphic, we may suppose that $d_n = 1$. Furthermore, by Corollary 1.4, we may assume that $n = \frac{d_1^2}{4} + d_1$, so $n = x^2 + 2x$.

Now, as in the proof of Theorem 1.2, by the Erdős–Gallai Theorem, there exists k with $1 \leq k \leq n$, such that

$$\sum_{i=1}^{k} d_i > k(k-1) + \sum_{i=k+1}^{n} \min\{k, d_i\}. \tag{2.2}$$

For each i with $1 \leq i \leq k$, replace d_i by d_1; the left hand side of (2.2) is not decreased, while the right hand side of (2.2) is unchanged, so (2.2) still holds. For each i with $k+1 \leq i \leq n$, replace d_i by 1; the left hand side of (2.2) is unchanged, while the right hand side of (2.2) has not increased, so (2.2) again holds. Then (2.2) reads $kd_1 > k(k-1) + (n-k)$, and consequently, rearranging terms, $(k-x-1)^2 - 1 < 0$. Thus $k = x+1$. Notice that for $1 \leq i \leq k$, if any of the original terms d_i had been less than d_1, we would have obtained $(k-x-1)^2 < 0$, which is impossible. Similarly, for $k+1 \leq i \leq n$, all the original terms d_i must have been all equal to one. Thus $\underline{d} = (d_1^k, 1^{n-k}) = ((2x)^{x+1}, 1^{x^2+x-1})$. So $\underline{d}$ has sum $2x(x+1) + x^2 + x - 1 = 3x^2 + 3x - 1$, which is odd, regardless of whether x is even or odd. This contradicts the hypothesis.

Now consider the case where d_1 is odd, say $d_1 = 2x - 1$. The theorem is trivial for $\underline{d} = (1^n)$, so we may assume that $x > 1$. We use essentially the same approach as we used in the even case, but the odd case is somewhat more complicated. By Corollary 1.4, assuming $\underline{d}$ is not graphic, we have $\frac{d_1^2}{4} + d_1 + 1 > n$, and hence, as d_1 is odd, $\frac{d_1^2}{4} + d_1 + \frac{3}{4} \geq n$. Thus, since $n \geq \left\lfloor \frac{d_1^2}{4} + d_1 \right\rfloor = \frac{d_1^2}{4} + d_1 - \frac{1}{4}$, we have $n = \frac{d_1^2}{4} + d_1 + \frac{3}{4}$ or $n = \frac{d_1^2}{4} + d_1 - \frac{1}{4}$. Thus there are two cases:

(i) $n = x^2 + x - 1$,

(ii) $n = x^2 + x$.

By the Erdős–Gallai Theorem, there exists k with $1 \leq k \leq n$, such that

$$\sum_{i=1}^{k} d_i > k(k-1) + \sum_{i=k+1}^{n} \min\{k, d_i\}. \tag{2.3}$$

As before, for each i with $1 \leq i \leq k$, replace d_i by d_1 and for each i with $k+1 \leq i \leq n$, replace d_i by d_n, and note that (2.3) again holds. Arguing as in the proof of Theorem 1.2, notice that if $k < d_n$, then (2.3) gives $kd_1 > k(k-1) + (n-k)k = k(n-1)$, and so $d_1 \geq n$. In both cases (i) and (ii) we would have $2x-1 \geq n \geq x^2+x-1$ and hence $x \leq 1$, contrary to our assumption. Thus $k \geq d_n$ and (2.3) reads $kd_1 > k(k-1) + (n-k)d_n$, and consequently, rearranging terms, we obtain in the respective cases:

(i) $d_n x^2 - d_n k + k^2 + d_n x - 2kx - d_n < 0.$

(ii) $d_n x^2 - d_n k + k^2 + d_n x - 2kx < 0,$

In both cases we have $d_n x^2 - d_n k + k^2 + d_n x - 2kx - d_n < 0$. Consider $d_n x^2 - d_n k + k^2 + d_n x - 2kx - d_n$ as a quadratic in k. For this to be negative, its discriminant, $4d_n + d_n^2 + 4x^2 - 4d_n x^2$, must be positive. If $d_n > 1$ we obtain $x^2 < \frac{4d_n + d_n^2}{4(d_n-1)}$. For $d_n = 2$ we have $x^2 < 3$ and so $x = 1$, contrary to our assumption. Similarly, for $d_n = 3$ we have $x^2 < \frac{21}{8}$ and so again $x = 1$. For $d_n \geq 4$, the function $\frac{4d_n + d_n^2}{4(d_n-1)}$ is monotonic increasing in d_n. So, as $d_n \leq d_1$,

$$x^2 < \frac{4d_1 + d_1^2}{4(d_1-1)} = \frac{4x^2+4x-3}{8x-8} < \frac{x^2+x}{2(x-1)},$$

which again gives $x = 1$. We conclude that $d_n = 1$.

So the two cases are:

(i) $x^2 - k + k^2 + x - 2kx - 1 = (k-x)(k-x-1) - 1 < 0.$

(ii) $x^2 - k + k^2 + x - 2kx = (k-x)(k-x-1) < 0,$

In case (ii) we must have $x < k < x+1$, but this is impossible for integer k and x.

In case (i), either $k = x$ or $k = x+1$. Notice that for $1 \leq i \leq k$, if any of the original terms d_i had been less than d_1, we would have obtained $(k-x)(k-x-1) < 0$, which is impossible. Similarly, for $k+1 \leq i \leq n$, all the original terms d_i must have been all equal to one. Thus $\underline{d} = (d_1^k, 1^{n-k})$. Consequently, if $k = x$, we have $\underline{d} = ((2x-1)^x, 1^{x^2-1})$ as $n = x^2 + x - 1$. In this case, $\underline{d}$ has sum $x(2x-1) + x^2 - 1 = 3x^2 - x - 1$, which is odd, regardless of whether x is even or odd, contradicting the hypothesis. On the other hand, if $k = x+1$, we have $\underline{d} = ((2x-1)^{x+1}, 1^{x^2-2})$. Here, $\underline{d}$ has sum $(2x-1)(x+1) + x^2 - 2 = 3x^2 + x - 3$, which is again odd, regardless of whether x is even or odd, contrary to the hypothesis. □

References

[1] M. D. Barrus, S. G. Hartke, K. F. Jao, and D. B. West, Length thresholds for graphic lists given fixed largest and smallest entries and bounded gaps, *Discrete Math.* **312** (2012), no. 9, 1494–1501.

[2] G. Cairns and S. Mendan, *Symmetric bipartite graphs and graphs with loops*, preprint.

[3] P. L. Hammer, T. Ibaraki, and B. Simeone, Threshold sequences, *SIAM J. Algebraic Discrete Methods* **2** (1981), no. 1, 39–49.

[4] S.Y.R. Li, Graphic sequences with unique realization, *J. Combinatorial Theory Ser. B* **19** (1975), no. 1, 42–68.

[5] A. Tripathi and H. Tyagi, A simple criterion on degree sequences of graphs, *Discrete Appl. Math.* **156** (2008), no. 18, 3513–3517.

[6] A. Tripathi, S. Venugopalan and D. B. West, A short constructive proof of the Erdős-Gallai characterization of graphic lists, *Discrete Math.* **310** (2010), no. 4, 843–844.

[7] J.-H. Yin, Conditions for r-graphic sequences to be potentially $K_{m+1}^{(r)}$-graphic, *Discrete Math.* **309** (2009), no. 21, 6271–6276.

[8] I. È. Zverovich and V. È. Zverovich, Contributions to the theory of graphic sequences, *Discrete Math.* **105** (1992), no. 1-3, 293–303.

ARS MATHEMATICA CONTEMPORANEA

Also available at http://amc-journal.eu
ISSN 1855-3966 (printed edn.), ISSN 1855-3974 (electronic edn.)
ARS MATHEMATICA CONTEMPORANEA 10 (2016) 85–90

Nowhere-zero 3-flows in graphs admitting solvable arc-transitive groups of automorphisms*

Xiangwen Li †
Department of Mathematics, Huazhong Normal University, Wuhan 430079, China

Sanming Zhou ‡
School of Mathematics and Statistics, The University of Melbourne, Parkville, VIC 3010, Australia

Received 18 February 2014, accepted 20 May 2014, published online 8 July 2015

Abstract

Tutte's 3-flow conjecture asserts that every 4-edge-connected graph has a nowhere-zero 3-flow. In this note we prove that every regular graph of valency at least four admitting a solvable arc-transitive group of automorphisms admits a nowhere-zero 3-flow.

Keywords: Integer flow, nowhere-zero 3-flow, vertex-transitive graph, arc-transitive graph, solvable group.

Math. Subj. Class.: 05C21, 05C25

1 Introduction

All graphs in this paper are finite and undirected, and all groups considered are finite. Let $\Gamma = (V(\Gamma), E(\Gamma))$ be a graph endowed with an orientation. For an integer $k \geq 2$, a *k-flow* [3] in Γ is an integer-valued function $f : E(\Gamma) \to \{0, \pm 1, \pm 2, \dots, \pm(k-1)\}$ such that, for every $v \in V(\Gamma)$,

$$\sum_{e \in E^+(v)} f(e) = \sum_{e \in E^-(v)} f(e),$$

where $E^+(v)$ is the set of edges of Γ with tail v and $E^-(v)$ the set of edges of Γ with head v. A k-flow f in Γ is called a *nowhere-zero k-flow* if $f(e) \neq 0$ for every $e \in E(\Gamma)$. Obviously,

*The authors are grateful to an anonymous referee for his/her comments which led to improved presentation.
†Li was supported by the National Natural Science Foundation of China (11171129).
‡Zhou was supported by a Future Fellowship of the Australian Research Council. Part of the work was done when he visited Huazhong Normal University in 2011.
E-mail addresses: xwli68@mail.ccnu.edu.cn (Xiangwen Li), smzhou@ms.unimelb.edu.au (Sanming Zhou)

if Γ admits a nowhere-zero k-flow, then Γ admits a nowhere-zero $(k+1)$-flow. It is also easy to see that whether a graph admits a nowhere-zero k-flow is independent of its orientation. The notion of nowhere-zero flows was introduced by Tutte in [20, 21] who proved that a planar graph admits a nowhere-zero 4-flow if and only if the Four Color Conjecture holds. The reader is referred to Jaeger [9] and Zhang [25] for surveys on nowhere-zero flows and to [3, Chapter 21] for an introduction to this area.

In [20, 21] Tutte proposed three celebrated conjectures on integer flows which are still open in general. One of them is the following well-known 3-flow conjecture (see e.g. [3, Conjecture 21.16]).

Conjecture 1.1. *(Tutte's 3-flow conjecture) Every 4-edge-connected graph admits a nowhere-zero 3-flow.*

This conjecture has been extensively studied in over four decades; see e.g. [6, 7, 10, 11, 12, 19, 23, 24]. Solving a long-standing conjecture by Jaeger [8] (namely the weak 3-flow conjecture), recently Thomassen [19] proved that every 8-edge-connected graph admits a nowhere-zero 3-flow. This breakthrough was further improved by Lovász, Thomassen, Wu and Zhang [12] who proved the following result.

Theorem 1.2. *([12, Theorem 1.7]) Every 6-edge-connected graph admits a nowhere-zero 3-flow.*

It is well known [22] that every vertex-transitive graph of valency $d \geq 1$ is d-edge-connected. Thus, when restricted to the class of vertex-transitive graphs, Conjecture 1.1 asserts that every vertex-transitive graph of valency at least four admits a nowhere-zero 3-flow. Due to Theorem 1.2 this is now boiled down to vertex-transitive graphs of valency 5, since every regular graph with even valency admits a nowhere-zero 2-flow. In an attempt to Tutte's 3-flow conjecture for Cayley graphs, Potočnik, Škoviera and Škerkovski [16] proved the following result. (It is well known [2] that every Cayley graph is vertex-transitive, but the converse is not true.)

Theorem 1.3. *([16, Theorem 1.1]) Every Cayley graph of valency at least four on an abelian group admits a nowhere-zero 3-flow.*

This was generalized by Nánásiová and Škoviera [13] to Cayley graphs on nilpotent groups.

Theorem 1.4. *([13, Theorem 4.3]) Every Cayley graph of valency at least four on a nilpotent group admits a nowhere-zero 3-flow.*

It would be nice if one can prove Tutte's 3-flow conjecture for all vertex-transitive graphs. As a step towards this, we prove the following result in the present paper.

Theorem 1.5. *Every regular graph of valency at least four admitting a solvable arc-transitive group of automorphisms admits a nowhere-zero 3-flow.*

Note that any G-arc-transitive graph is necessarily G-vertex-transitive and G-edge-transitive. On the other hand, any G-vertex-transitive and G-edge-transitive graph with odd valency is G-arc-transitive. (This result is due to Tutte, and its combinatorial proof given in [4, Proposition 1.2] can be easily extended from $G = \mathrm{Aut}(\Gamma)$ to a subgroup G of $\mathrm{Aut}(\Gamma)$.) Therefore, Theorem 1.5 is equivalent to the following: For any solvable group

G, every G-vertex-transitive and G-edge-transitive graph with valency at least four admits a nowhere-zero 3-flow.

In Section 3 we will prove a weaker version (Claim 1) of Theorem 1.5, which together with Theorem 1.2 implies Theorem 1.5. Note that none of Theorems 1.5 and 1.4 is implied by the other, because not every Cayley graph is arc-transitive, and on the other hand an arc-transitive graph is not necessarily a Cayley graph (see e.g. [2, 17]).

At this point we would like to mention a few related results. Alspach and Zhang (1992) conjectured that every Cayley graph with valency at least two admits a nowhere-zero 4-flow. Since every 4-edge-connected graph admits a nowhere-zero 4-flow [8], this conjecture is reduced to the cubic case. Alspach, Liu and Zhang [1, Theorem 2.2] confirmed this conjecture for cubic Cayley graphs on solvable groups. This was then improved by Nedela and Škoviera [14] who proved that any counterexample to the conjecture of Alspach and Zhang must be a regular cover over a Cayley graph on an almost simple group. (A group G is *almost simple* if it satisfies $T \leq G \leq \mathrm{Aut}(T)$ for some simple group T.) In [15], Potočnik proved that every connected cubic graph that admits a solvable vertex-transitive group of automorphisms is 3-edge-colourable or isomorphic to the Petersen graph. This is another generalization of the result of Alspach, Liu and Zhang above, because Petersen graph is not a Cayley graph, and for cubic graphs 3-edge-colourability is equivalent to the existence of a nowhere-zero 4-flow.

It would be pleasing if one can replace arc-transitivity by vertex-transitivity in Theorem 1.5. As an intermediate step towards this, one may try to prove that every Cayley graph of valency at least four on a solvable group admits a nowhere-zero 3-flow, thus generalizing Theorem 1.4 and the above-mentioned result of Alspach, Liu and Zhang simultaneously.

2 Preparations

We follow [3] and [5, 18] respectively for graph- and group-theoretic terminology and notation. The derived subgroup of a group G is defined as $G' := [G, G]$, the subgroup of G generated by all commutators $x^{-1}y^{-1}xy$, $x, y \in G$. Define $G^{(0)} := G, G^{(1)} := G'$ and $G^{(i)} := (G^{(i-1)})'$ for $i \geq 1$. A group G is *solvable* if there exists an integer $n \geq 0$ such that $G^{(n)} = 1$; in this case the least integer n with $G^{(n)} = 1$ is called the *derived length* of G. Solvable groups with derived length 1 are precisely nontrivial abelian groups. In the proof of Theorem 1.5 we will use the fact that any solvable group contains an abelian normal subgroup with respect to which the quotient group has a smaller derived length.

All definitions in the next three paragraphs are standard and can be found in [2, Part Three] or [17].

Let G be a group acting on a set Ω. That is, for each $(\alpha, g) \in \Omega \times G$, there corresponds an element $\alpha^g \in \Omega$ such that $\alpha^1 = \alpha$ and $(\alpha^g)^h = \alpha^{gh}$ for any $\alpha \in \Omega$ and $g, h \in G$, where 1 is the identity element of G. We say that G is *transitive* on Ω if for any $\alpha, \beta \in \Omega$ there exists at least one element $g \in G$ such that $\alpha^g = \beta$, and *regular* if for any $\alpha, \beta \in \Omega$ there exists exactly one element $g \in G$ such that $\alpha^g = \beta$. The group G is *intransitive* on Ω if it is not transitive on Ω. A partition $\mathcal{P}$ of Ω is G-*invariant* if $P^g := \{\alpha^g : g \in G\} \in \mathcal{P}$ for any $P \in \mathcal{P}$ and $g \in G$, and *nontrivial* if $1 < |P| < |\Omega|$ for every $P \in \mathcal{P}$.

Suppose that Γ is a graph admitting G as a group of automophisms. That is, G acts on $V(\Gamma)$ (not necessarily faithfully) such that, for any $\alpha, \beta \in V(\Gamma)$ and $g \in G$, α and β are adjacent in Γ if and only if α^g and β^g are adjacent in Γ. (If K is the *kernel* of the action of G on $V(\Gamma)$, namely, the subgroup of all elements of G that fix every vertex of Γ, then

G/K is isomorphic to a subgroup of the automorphism group $\mathrm{Aut}(\Gamma)$ of Γ.) We say that Γ is *G-vertex-transitive* if G is transitive on $V(\Gamma)$, and *G-edge-transitive* if G is transitive on the set of edges of Γ. If Γ is G-vertex-transitive such that G is also transitive on the set of arcs of Γ, then Γ is called *G-arc-transitive*, where an *arc* is an ordered pair of adjacent vertices.

Let Γ be a graph and $\mathcal{P}$ a partition of $V(\Gamma)$. The *quotient graph* of Γ with respect to $\mathcal{P}$, denoted by $\Gamma_{\mathcal{P}}$, is the graph with vertex set $\mathcal{P}$ in which $P, Q \in \mathcal{P}$ are adjacent if and only if there exists at least one edge of Γ joining a vertex of P and a vertex of Q. For blocks $P, Q \in \mathcal{P}$ adjacent in $\Gamma_{\mathcal{P}}$, denote by $\Gamma[P, Q]$ the bipartite subgraph of Γ with vertex set $P \cup Q$ whose edges are those of Γ between P and Q. In the case when all blocks of $\mathcal{P}$ are independent sets of Γ and $\Gamma[P, Q]$ is a t-regular bipartite graph for each pair of adjacent $P, Q \in \mathcal{P}$, where $t \geq 1$ is an integer independent of (P, Q), we say that Γ is a *multicover* of $\Gamma_{\mathcal{P}}$. A multicover with $t = 1$ is thus a topological cover in the usual sense. In the proof of Theorem 1.5, we will use the following lemma in the case when $k = 3$.

Lemma 2.1. *Let $k \geq 2$ be an integer. If a graph admits a nowhere-zero k-flow, then its multicovers all admit a nowhere-zero k-flow.*

Proof. Using the notation above, let Γ be a multicover of $\Sigma := \Gamma_{\mathcal{P}}$. Suppose that Σ admits a nowhere-zero k-flow f (with respect to some orientation). For each oriented edge (P, Q) of Σ, orient the edges of the t-regular bipartite graph $\Gamma[P, Q]$ in such a way that they all have tails in P and heads in Q, and then assign $f(P, Q)$ to each of them. Denote this nowhere-zero function on the oriented edges of Γ by g, and denote the oriented edge of Γ with tail α and head β by (α, β). It can be verified that, for any $P \in \mathcal{P}$ and $\alpha \in P$, $\sum_{(\alpha,\beta)\in E^+_\Gamma(\alpha)} g(\alpha, \beta) = \sum_{(P,Q)\in E^+_\Sigma(P)} t \cdot f(P, Q)$ and $\sum_{(\beta,\alpha)\in E^-_\Gamma(\alpha)} g(\beta, \alpha) = \sum_{(Q,P)\in E^-_\Sigma(P)} t \cdot f(Q, P)$. Since f is a nowhere-zero k-flow in Σ, for every $P \in \mathcal{P}$, we have

$$\sum_{(P,Q)\in E^+_\Sigma(P)} f(P, Q) = \sum_{(Q,P)\in E^-_\Sigma(P)} f(Q, P).$$

Therefore, for every $\alpha \in V(\Gamma)$, we have

$$\sum_{(\alpha,\beta)\in E^+_\Gamma(\alpha)} g(\alpha, \beta) = \sum_{(\beta,\alpha)\in E^-_\Gamma(\alpha)} g(\beta, \alpha)$$

and so g is a nowhere-zero k-flow in Γ. □

If Γ is a G-vertex-transitive graph, then for any normal subgroup N of G, the set $\mathcal{P}_N := \{\alpha^N : \alpha \in V(\Gamma)\}$ of N-orbits on $V(\Gamma)$ is a G-invariant partition of $V(\Gamma)$, called a *G-normal partition* of $V(\Gamma)$ [17], where $\alpha^N := \{\alpha^g : g \in N\}$. Denote the corresponding quotient graph by $\Gamma_N := \Gamma_{\mathcal{P}_N}$. The quotient group G/N induces an action on $\mathcal{P}_N$ defined by $(\alpha^N)^{Ng} = (\alpha^g)^N$. The following observations can be easily proved (see e.g. [17]).

Lemma 2.2. *([17]) Let Γ be a connected G-vertex-transitive graph, and N a normal subgroup of G that is intransitive on $V(\Gamma)$. Then the following hold:*

(a) *Γ_N is G/N-vertex-transitive under the induced action of G/N on $\mathcal{P}_N$;*
(b) *for $P, Q \in \mathcal{P}_N$ adjacent in Γ_N, $\Gamma[P, Q]$ is a regular subgraph of Γ;*
(c) *if in addition Γ is G-arc-transitive, then Γ_N is G/N-arc-transitive and Γ is a multicover of Γ_N.*

3 Proof of Theorem 1.5

By Theorem 1.2, in order to prove Theorem 1.5 it suffices to prove that, for any finite solvable group G, every G-arc-transitive graph of valency five admits a nowhere-zero 3-flow. We will prove the following seemingly stronger but equivalent result:

Claim 1. For any solvable group G, every G-arc-transitive graph with valency at least four and not divisible by three admits a nowhere-zero 3-flow.

We will prove this by induction on the derived length of the solvable group G, using Theorem 1.3 as the base case. We will use the following fact [2, Lemma 16.3]: a graph is isomorphic to a Cayley graph if and only if its automorphism group contains a subgroup that is regular on the vertex set. Denote by $\mathrm{val}(\Gamma)$ the valency of a regular graph Γ.

Without loss of generality we may assume that the solvable group G is faithful on the vertex set of the graph under consideration for otherwise we can replace G by its quotient group (which is also solvable) by the kernel of G on the vertex set. Under this assumption G is isomorphic to a subgroup of the automorphism group of the graph. We may also assume that the graph under consideration is connected (for otherwise we consider its components). We make induction on the derived length $n(G)$ of G.

Suppose that $n(G) = 1$ and Γ is a G-arc-transitive graph with $\mathrm{val}(\Gamma) \geq 4$. Then G is abelian and so is regular on $V(\Gamma)$. (A transitive abelian group must be regular.) Since G is isomorphic to a subgroup of $\mathrm{Aut}(\Gamma)$, it follows that Γ is isomorphic to a Cayley graph on G. Thus, by Theorem 1.3, Γ admits a nowhere-zero 3-flow.

Assume that, for some integer $n \geq 1$, the result (in Claim 1) holds for any solvable group of derived length at most n. Let G be a solvable group with derived length $n(G) = n + 1$. Let Γ be a connected G-arc-transitive graph such that $\mathrm{val}(\Gamma) \geq 4$ and $\mathrm{val}(\Gamma)$ is not divisible by 3. If $\mathrm{val}(\Gamma)$ is even, then Γ admits a nowhere-zero 2-flow and hence a nowhere-zero 3-flow. So we assume that $\mathrm{val}(\Gamma) \geq 5$ is odd. Since 3 does not divide $\mathrm{val}(\Gamma)$ by our assumption, every prime factor of $\mathrm{val}(\Gamma)$ is no less than 5. Since G is solvable, it contains an abelian normal subgroup N such that the quotient group G/N has derived length at most $n(G) - 1 = n$. Note that G/N is solvable (as any quotient group of a solvable group is solvable) and $N \neq 1$ (for otherwise $G/N \cong G$ would have derived length $n(G)$). If N is transitive on $V(\Gamma)$, then it is regular on $V(\Gamma)$ as N is abelian. In this case Γ is isomorphic to a Cayley graph on N and so admits a nowhere-zero 3-flow by Theorem 1.3.

In what follows we assume that N is intransitive on $V(\Gamma)$. By Lemma 2.2, Γ_N is a connected G/N-arc-transitive graph, and Γ is a multicover of Γ_N. Thus $\mathrm{val}(\Gamma_N)$ is a divisor of $\mathrm{val}(\Gamma)$ and so is not divisible by 3. If $\mathrm{val}(\Gamma_N) = 1$, then Γ is a regular bipartite graph of valency at least two and so admits a nowhere-zero 3-flow [3]. Assume that $\mathrm{val}(\Gamma_N) > 1$. Then $\mathrm{val}(\Gamma_N) \geq 5$ and every prime factor of $\mathrm{val}(\Gamma_N)$ is no less than 5. Thus, since G/N is solvable of derived length at most n, by the induction hypothesis, Γ_N admits a nowhere-zero 3-flow. Since Γ is a multicover of Γ_N, by Lemma 2.1, Γ admits a nowhere-zero 3-flow. This completes the proof of Claim 1 and hence the proof of Theorem 1.5.

References

[1] B. Alspach, Y.-P. Liu and C.-Q. Zhang, Nowhere-zero 4-flows and Cayley graphs on solvable groups, *SIAM J. Discrete Math.* **9** (1996), 151–154.

[2] N. L. Biggs, *Algebraic Graph Theory*, 2nd ed., Cambridge University Press, Cambridge, 1993.

[3] J. A. Bondy, U. S. R. Murty, *Graph Theory*, Springer, 2008.

[4] Y. Cheng and J. Oxley, On weakly symmetric graphs of order twice a prime, *J. Combin. Theory Ser. B* **42** (1987), 196–211.

[5] J. D. Dixon and B. Mortimer, *Permutation Groups*, Springer, New York, 1996.

[6] G. Fan and C. Zhou, Degree sum and Nowhere-zero 3-flows, *Discrete Math.* **308** (2008), 6233–6240.

[7] G. Fan and C. Zhou, Ore condition and Nowhere-zero 3-flows, *SIAM J. Discrete Math.* **22** (2008), 288–294.

[8] F. Jaeger, Flows and generalized coloring theorems in graphs, *J. Combin. Theory Ser. B* **26** (1979), 205–216.

[9] F. Jaeger, Nowhere-zero flow problems, in: L. W. Beineke and R. J. Wilson (eds.), *Selected Topics in Graph Theory*, Vol. 3, Academic Press, San Diego, 1988, 71–95.

[10] H. -J. Lai, Nowhere-zero 3-flows in locally connected graphs, *J. Graph Theory* **4** (2003), 211–219.

[11] X. Li, Y. Shao, H.-J. Lai, Degree condition and Z_3-connectivity, *Discrete Math.* **312** (2012), 1658–1669.

[12] L. M. Lovász, C. Thomassen, Y. Wu, C. -Q. Zhang, Nowhere-zero 3-flows and modulo k-orientations, *J. Combin. Theory Ser. B* **103** (2013), 587–598.

[13] M. Nánásiová and M. Škoviera, Nowhere-zero flows in Cayley graphs and Sylow 2-subgroups, *J. Alg. Combin.* **30** (2009), 103–110.

[14] R. Nedela and M. Škoviera, Cayley snarks and almost simple groups, *Combinatorica* **21** (2001), 583–590.

[15] P. Potočnik, Edge-colourings of cubic graphs admitting a solvable vertex-transitive group of automorphisms, *J. Combin. Theory Ser. B* **91** (2004), 289–300.

[16] P. Potočnik, M. Škoviera and R. Škrekovski, Nowhere-zero 3-flows in abelian Cayley graphs, *Discrete Math.* **297** (2005), 119–127.

[17] C. E. Praeger, Finite transitive permutation groups and finite vertex transitive graphs, in: G. Hahn and G. Sabidussi (eds.), *Graph Symmetry*, Kluwer Academic Publishing, Dordrecht, 1997, 277–318.

[18] J. J. Rotman, *An Introduction to the Theory of Groups*, 4th ed., Springer, New York, 1995.

[19] C. Thomassen, The weak 3-flow conjecture and the weak circular flow conjecture, *J. Combin. Theory Ser. B* **102** (2012), 521–529.

[20] W. T. Tutte, A contribution on the theory of chromatic polynomial, *Canad. J. Math.* **6** (1954), 80–91.

[21] W. T. Tutte, On the algebraic theory of graph colorings *J. Combin. Theory* **1** (1966), 15–50.

[22] A. E. Watkins, Connectivity of transitive graphs, *J. Combin. Theory* **8** (1970), 23–29.

[23] J. Yan, Nowhere-zero 3-flows and Z_3-connectivity of a family of graphs, *Discrete Math.* **311** (2011), 1988–1994.

[24] F. Yang and X. Li, Nowhere-zero 3-flows in dihedral Cayley graphs, *Information Processing Letters* **111** (2011), 416–419.

[25] C.-Q. Zhang, *Integer Flows and Cycle Covers of Graphs*, Marcel Dekker Inc., New York, 1997.

ARS MATHEMATICA CONTEMPORANEA

Also available at http://amc-journal.eu
ISSN 1855-3966 (printed edn.), ISSN 1855-3974 (electronic edn.)
ARS MATHEMATICA CONTEMPORANEA 10 (2016) 91–98

A note on nowhere-zero 3-flow and Z_3-connectivity

Fuyuan Chen

Center for Discrete Mathematics, Fuzhou University, Fuzhou, P. R. China,
Institute of Statistics and Applied Mathematics, Anhui University of Finance and Economics, Bengbu, 233030, P. R. China

Bo Ning

Department of Applied Mathematics, School of Science, Northwestern Polytechnical University, Xi'an, P. R. China,
Center for Applied Mathematics, Tianjin University, 300072, P. R. China

Received 22 May 2014, accepted 14 July 2014, published online 8 July 2015

Abstract

There are many major open problems in integer flow theory, such as Tutte's 3-flow conjecture that every 4-edge-connected graph admits a nowhere-zero 3-flow, Jaeger et al.'s conjecture that every 5-edge-connected graph is Z_3-connected and Kochol's conjecture that every bridgeless graph with at most three 3-edge-cuts admits a nowhere-zero 3-flow (an equivalent version of 3-flow conjecture). Thomassen proved that every 8-edge-connected graph is Z_3-connected and therefore admits a nowhere-zero 3-flow. Furthermore, Lovász, Thomassen, Wu and Zhang improved Thomassen's result to 6-edge-connected graphs. In this paper, we prove that: (1) Every 4-edge-connected graph with at most seven 5-edge-cuts admits a nowhere-zero 3-flow. (2) Every bridgeless graph containing no 5-edge-cuts but at most three 3-edge-cuts admits a nowhere-zero 3-flow. (3) Every 5-edge-connected graph with at most five 5-edge-cuts is Z_3-connected. Our main theorems are partial results to Tutte's 3-flow conjecture, Kochol's conjecture and Jaeger et al.'s conjecture, respectively.

Keywords: Integer flow, nowhere-zero 3-flow, Z_3-connected, modulo 3-orientation, edge-cuts.

Math. Subj. Class.: 05C21, 05C40

E-mail addresses: chenfuyuan19871010@163.com (Fuyuan Chen), ningbo-math84@mail.nwpu.edu.cn (Bo Ning)

1 Introduction

All graphs considered in this paper are loopless, but allowed to have multiple edges. A graph G is called *k-edge-connected*, if $G-S$ is connected for each edge set S with $|S| < k$. Let X, Y be two disjoint subsets of $V(G)$. Let $\partial_G(X, Y)$ be the set of edges of G with one end in X and the other in Y. In particular, if $Y = \overline{X}$, we simply write $\partial_G(X)$ for $\partial_G(X, Y)$, which is the *edge-cut* of G associated with X. The edge set $C = \partial_G(X)$ is called a *k-edge-cut* if $|\partial_G(X)| = k$. If X is nontrivial, we use G/X to denote the graph obtained from G by replacing X by a single vertex x that is incident with all the edges in $\partial_G(X)$.

Let D be an orientation of $E(G)$. The *out-cut* of D associated with X, denoted by $\partial^+_D(X)$, is the set of arcs of D whose tails lie in X. Analogously, the *in-cut* of D associated with X, denoted by $\partial^-_D(X)$, is the set of arcs of D whose heads lie in X. We refer to $|\partial^+_D(X)|$ and $|\partial^-_D(X)|$ as the out-degree and in-degree of X, and denote these quantities by $d^+_D(X)$ and $d^-_D(X)$, respectively.

Definition 1.1. (1) An orientation D of $E(G)$ is called a *modulo 3-orientation* if

$$d^+_D(v) - d^-_D(v) \equiv 0 \pmod 3$$

for every vertex $v \in V(G)$.

(2) A pair (D, f) is called a *nowhere-zero 3-flow* of G if D is an orientation of $E(G)$ and f is a function from $E(G)$ to $\{\pm 1, \pm 2\}$, such that

$$\sum_{e \in \partial^+_D(v)} f(e) = \sum_{e \in \partial^-_D(v)} f(e)$$

for every vertex $v \in V(G)$.

The 3-flow conjecture, proposed by Tutte as a dual version of Grötzsch's 3-color theorem for planar graphs, may be one of the most major open problems in integer flow theory.

Conjecture 1.2 (3-Flow conjecture, Tutte [9])**.** *Every 4-edge-connected graph admits a nowhere-zero 3-flow.*

Kochol proved that Tutte's 3-flow conjecture is equivalent to the following two conjectures.

Conjecture 1.3 (Kochol [4])**.** *Every 5-edge-connected graph admits a nowhere-zero 3-flow.*

Conjecture 1.4 (Kochol [5])**.** *Every bridgeless graph with at most three 3-edge-cuts admits a nowhere-zero 3-flow.*

A weakened version of Conjecture 1.2, the so-called weak 3-flow conjecture, was proposed by Jaeger.

Conjecture 1.5 (Weak 3-flow conjecture, Jaeger [2])**.** *There is a natural number h such that every h-edge-connected graph admits a nowhere-zero 3-flow.*

Lai and Zhang [6] and Alon et al. [1] gave partial results on Conjectures 1.2 and 1.5.

Theorem 1.6 (Lai and Zhang [6]). *Every $4\lceil \log_2 n_0 \rceil$-edge-connected graph with at most n_0 odd-degree vertices admits a nowhere-zero 3-flow.*

Theorem 1.7 (Alon, Linial and Meshulam [1]). *Every $2\lceil \log_2 n \rceil$-edge-connected graph with n vertices admits a nowhere-zero 3-flow.*

Recently, Thomassen [8] confirmed weak 3-flow conjecture. He proved

Theorem 1.8 (Thomassen [8]). *Every 8-edge-connected graph is Z_3-connected and therefore admits a nowhere-zero 3-flow.*

Thomassen's method was further refined by Lovász, Thomassen, Wu and Zhang [7] to obtain the following theorem.

Theorem 1.9 (Lovász, Thomassen, Wu and Zhang [7]). *Every 6-edge-connected graph is Z_3-connected and therefore admits a nowhere-zero 3-flow.*

For more results on Tutte's 3-flow conjecture, we refer the reader to the introduction part of [7] and the book written by Zhang [11].

In this paper, we will give the following conjecture which is equivalent to Tutte's 3-flow conjecture.

Conjecture 1.10. *Every 5-edge-connected graph with minimum degree at least 6 has a nowhere-zero 3-flow.*

To prove the equivalence of Conjectures 1.2 and 1.10, the following lemma is needed.

Lemma 1.11 (Tutte [10]). *Let $F(G,k)$ be the number of nowhere-zero k-flows of G. Then $F(G,k) = F(G/e,k) - F(G \setminus e,k)$ if e is not a loop of G.*

Proposition 1.12. *Conjectures 1.2 and 1.10 are equivalent.*

Proof. It is obvious that Conjecture 1.2 implies Conjecture 1.3, and Conjecture 1.3 implies Conjecture 1.10. Now we prove that Conjecture 1.10 can imply Conjecture 1.3. Let G be a 5-edge-connected graph. Let G' be the graph obtained from G by gluing $|V(G)|$ disjoint copies of K_7, such that for each such copy H_i, $|V(H_i) \cap V(G)| = 1$ $(i = 1, 2, \cdots, |V(G)|)$. Then G' is 5-edge-connected and its minimum degree is at least 6, and thus has a nowhere-zero 3-flow. By Lemma 1.11, G has a nowhere-zero 3-flow. Therefore Conjecture 1.10 implies Conjecture 1.3. Note that Conjecture 1.2 is equivalent to Conjecture 1.3. This completes the proof. □

Our first main result is the following theorem.

Theorem 1.13. *Let G be a bridgeless graph and let $P = \{C = \partial_G(X) : |C| = 3,\ X \subset V(G)\}$ and $Q = \{C = \partial_G(X) : |C| = 5,\ X \subset V(G)\}$. If $2|P| + |Q| \leq 7$, then G has a modulo 3-orientation (and therefore has a nowhere-zero 3-flow).*

As corollaries of Theorem 1.13, we obtain Theorems 1.14 and 1.15.

Theorem 1.14. *Every 4-edge-connected graph with at most seven 5-edge-cuts admits a nowhere-zero 3-flow.*

Theorem 1.15. *Every bridgeless graph containing no 5-edge-cuts but at most three 3-edge-cuts admits a nowhere-zero 3-flow.*

Remark. *The number of 3-edge-cuts in Theorem 1.15 can not be improved from three to four, since K_4 or any graph contractable to K_4 has no nowhere-zero 3-flow.*

Theorems 1.14 and 1.15 partially confirm Conjectures 1.2 and 1.4, respectively.

Definition 1.16. (1) A mapping $\beta_G : V(G) \mapsto Z_k$ is called a *Z_k-boundary* of G if

$$\sum_{v \in V(G)} \beta_G(v) \equiv 0 \pmod k$$

(2) A graph G is called *Z_k-connected*, if for every Z_k-boundary β_G, there is an orientation D_{β_G} and a function f_{β_G}: $E(G) \mapsto Z_k - \{0\}$, such that

$$\sum_{e \in \partial^+_{D_{\beta_G}}(v)} f_{\beta_G}(e) - \sum_{e \in \partial^-_{D_{\beta_G}}(v)} f_{\beta_G}(e) \equiv \beta_G(v) \pmod k$$

for every vertex $v \in V(G)$.

Jaeger, Linial, Payan and Tarsi [3] conjectured that

Conjecture 1.17 (Jaeger, Linial, Payan and Tarsi [3]). *Every 5-edge-connected graph is Z_3-connected.*

By applying a similar argument as in the proof of Theorem 1.13, we could obtain the second main result, which is a partial result to Conjecture 1.17.

Theorem 1.18. *Every 5-edge-connected graph with at most five 5-edge-cuts is Z_3-connected.*

In the next section, some necessary preliminaries will be given. In Sections 3 and 4, proofs of Theorems 1.13 and 1.18 will be given, respectively.

2 Preliminaries

In this section, we will give additional but necessary notations and definitions, and then give some useful lemmas.

Definition 2.1. Let β_G be a Z_3-boundary of G. An orientation D of G is called a *β_G-orientation* if

$$d^+_D(v) - d^-_D(v) \equiv \beta_G(v) \pmod 3$$

for every vertex $v \in V(G)$.

Let G be a graph and A be a vertex subset of G. The *degree* of A, denoted by $d_G(A)$, is the number of edges with precisely one end in A. Moreover if $A = \{x\}$, we simply write $d_G(x)$.

Let G be a graph and β_G be a Z_3-boundary of G. Define a mapping $\tau_G : V(G) \mapsto \{0, \pm 1, \pm 2, \pm 3\}$ such that, for each vertex $x \in V(G)$,

$$\tau_G(x) \equiv \begin{cases} \beta_G(x) & \pmod 3 \\ d_G(x) & \pmod 2. \end{cases}$$

Now, the mapping τ_G can be further extended to any nonempty vertex subset A as follows:

$$\tau_G(A) \equiv \begin{cases} \beta_G(A) & \pmod 3 \\ d_G(A) & \pmod 2. \end{cases}$$

where $\beta_G(A) \equiv \sum_{x\in A} \beta_G(x) \in \{0, 1, 2\} \pmod 3$.

Proposition 2.2. *Let G be a graph and A be a vertex subset of G.*
(1) If $d_G(A) \le 5$, then $d_G(A) \le 4 + |\tau_G(A)|$.
(2) If $d_G(A) \ge 6$, then $d_G(A) \ge 4 + |\tau_G(A)|$.

Proposition 2.2 follows from the fact that $|\tau_G(A)| \le 3$ and $d_G(A) - |\tau_G(A)|$ is even.

Lemma 2.3 (Tutte [9])**.** *Let G be a graph.*
(1) G has a nowhere-zero 3-flow if and only if G has a modulo 3-orientation.
(2) G has a nowhere-zero 3-flow if and only if G has a β_G-orientation with $\beta_G = 0$.

The following lemma is Theorem 3.1 in [7] by Lovász et al. This lemma will play the main role in our proofs.

Lemma 2.4 (Lovász, Thomassen, Wu and Zhang [7])**.** *Let G be a graph, β_G be a Z_3-boundary of G, and let $z_0 \in V(G)$ and D_{z_0} be a pre-orientation of $E(z_0)$ of all edges incident with z_0. Assume that*
(i) $|V(G)| \ge 3$.
(ii) $d_G(z_0) \le 4 + |\tau_G(z_0)|$ and $d^+_{D_{z_0}}(z_0) - d^-_{D_{z_0}}(z_0) \equiv \beta_G(z_0) \pmod 3$, and
(iii) $d_G(A) \ge 4 + |\tau_G(A)|$ for each nonempty vertex subset A not containing z_0 with $|V(G) \setminus A| > 1$.
Then the pre-orientation D_{z_0} of $E(z_0)$ can be extended to an orientation D of the entire graph G, that is, for every vertex x of G,

$$d^+_D(x) - d^-_D(x) \equiv \beta_G(x) \pmod 3.$$

3 Proof of Theorem 1.13

If not, suppose that G is a counterexample, such that $|V(G)| + |E(G)|$ is as small as possible. Let $P' = \{x \in V(G) : d_G(x) = 3\}$ and $Q' = \{x \in V(G) : d_G(x) = 5\}$.

Claim 3.1. $|V(G)| \ge 3$.

Proof. If $|V(G)| = 1$, then G has a nowhere-zero 3-flow, a contradiction. If $|V(G)| = 2$, let $V(G) = \{x, y\}$, then all the edges of G are all between x and y. Since G is bridgeless, $|E(G)| \ge 2$. Let a be the integer in $\{0, 1, 2\}$ such that $a \equiv |E(G)| - a \pmod 3$. Orient a edges from x to y and the remaining $|E(G)| - a$ edges from y to x. Clearly, the resulting orientation is a modulo 3-orientation of G, a contradiction. Therefore $|V(G)| \ge 3$. □

Claim 3.2. *G is 3-edge-connected, and G has no nontrivial 3-edge-cuts.*

Proof. If G has a vertex x of degree 2, then suppose that $xx_1, xx_2 \in E(G)$. By the minimality of G, $(G - \{xx_1, xx_2\}) \cup \{x_1x_2\}$ has a nowhere-zero 3-flow f'. However, f' can be extended to a nowhere-zero 3-flow f of G, a contradiction. If G has a nontrivial k-edge-cut ($k = 2, 3$), then contract one side and find a mod 3-orientation by the minimality of G. Merge such two mod 3-orientations and we will get one for G, a contradiction. □

Claim 3.3. *For any $U \subset V(G)$, if $d_G(U) \leq 5$ and $|U| \geq 2$, then $U \cap (P' \cup Q') \neq \emptyset$.*

Proof. If not, choose U to be a minimal one such that: for any $A \subset U$ with $2 \leq |A| < |U|$, we have $d_G(A) \geq 6$.

By the minimality of G, G/U has a modulo 3-orientation D' which is a partial modulo 3-orientation of G, such that $d^+_{D'}(x) \equiv d^-_{D'}(x) \pmod 3$ for each $x \in V(G) \setminus U$.

Let G' be a graph obtained from G by contracting $V(G) \setminus U$ as z_0 and let $\beta_{G'} = 0$.

(i) Since $V(G') = U + z_0$, $|V(G')| = |U| + 1 \geq 3$.

(ii) Since $d_{G'}(z_0) = d_G(U) \leq 5$, by Proposition 2.2 (1), $d_{G'}(z_0) \leq 4 + |\tau_{G'}(z_0)|$.

(iii) By the assumption and minimality of U, we have that for any $A \subset U$, $d_G(A) \neq 5$ and $d_G(A) \neq 3$. If $d_G(A) = 4$, then $d_{G'}(A) = d_G(A) = 4$ and $\tau_{G'}(A) = \beta_{G'}(A) = \beta_G(A) = 0$. Thus $d_{G'}(A) = 4 = 4 + |\tau_{G'}(A)|$. If $d_G(A) \geq 6$, then by Proposition 2.2 (2), $d_{G'}(A) = d_G(A) \geq 4 + |\tau_{G'}(A)|$.

By Lemma 2.4, we could see that the pre-orientation of $E'(z_0)$ of all edges incident with z_0 can be extended to a $\beta_{G'}$-orientation of G'. Then G has a modulo 3-orientation, which is a contradiction. □

Let G'_1 be a graph obtained from G by adding a new vertex z_0 and $2|P'| + |Q'|$ edges between z_0 and $P' \cup Q'$, such that:

(i) For each vertex $v \in P'$, we add two arcs with the same direction between it and z_0; and

(ii) For each vertex $v \in Q'$, we add one arc between it and z_0.

If $2|P'| + |Q'| \leq 5$, then all added arcs could be from z_0 to $P' \cup Q'$. Define $\beta_{G'_1}$ as follows:

(1) $\beta_{G'_1}(x) = 0$ if $x \notin (P' \cup Q') + z_0$;

(2) $\beta_{G'_1}(x) = 1$ if $x \in P'$;

(3) $\beta_{G'_1}(x) = 2$ if $x \in Q'$;

(4) $\beta_{G'_1}(z_0) \equiv 2|P'| + |Q'| \pmod 3$ and $\beta_{G'_1}(z_0) \in \{0, 1, 2\}$.

If $2|P'| + |Q'| = 6$ or 7, in this case, if $|P'| \neq 0$, choose one vertex $v \in P'$, such that the two arcs with ends z_0 and v are from v to z_0, the other arcs incident with z_0 are all directed from z_0. If $|P'| = 0$, then two arcs are from Q' to z_0, the others verse. Define $\beta_{G'_1}$ as follows:

(1) $\beta_{G'_1}(x) = 0$ if $x \notin (P' \cup Q') + z_0$;

(2) $\beta_{G'_1}(x) = 2$ if $x \in Q'$ and the arc (z_0, x) exists or $x \in P'$ and the two arcs with ends z_0 and x are from x to z_0;

(3) $\beta_{G'_1}(x) = 1$ if $x \in Q'$ and the arc (x, z_0) exists or $x \in P'$ and the two arcs with ends z_0 and x are from z_0 to x;

(4) $\beta_{G'_1}(z_0) \equiv (2|P'| + |Q'| - 2) - 2 \pmod 3$.

Now $d_{G'_1}(z_0) \leq 4 + |\tau_{G'_1}(z_0)|$ and $|V(G'_1)| = |V(G)| + 1 \geq 4$. We claim that: $d_{G'_1}(A) \geq 4 + |\tau_{G'_1}(A)|$, for each nonempty vertex subset A not containing z_0 with $|V(G'_1) \setminus A| > 1$.

If $A \cap (P' \cup Q') = \emptyset$, then by Claim 3.3, $d_G(A) = 4$ or $d_G(A) \geq 6$. In each case we could get that $d_{G'_1}(A) = d_G(A) \geq 4 + |\tau_{G'_1}(A)|$.

If $A \cap (P' \cup Q') \neq \emptyset$, then by Claim 3.2, $d_{G'_1}(A) \geq 5$. If $d_{G'_1}(A) = 5$, then $d_G(A) = 3$ or 4 and $|A \cap (P' \cup Q')| = 1$, and it follows that $\beta_{G'_1}(A) = 1$ or 2, and $|\tau_{G'_1}(A)| = 1$. Thus $d_{G'_1}(A) \geq 4 + |\tau_{G'_1}(A)|$. If $d_{G'_1}(A) \geq 6$, by Proposition 2.2 (2), we have that $d_{G'_1}(A) \geq 4 + |\tau_{G'_1}(A)|$.

Now G_1' satisfies all the conditions of Lemma 2.4. By Lemma 2.4, G_1' has a $\beta_{G_1'}$-orientation extended from the pre-orientation of $E_1'(z_0)$ of all edges incident with z_0, which implies that G has a β_G-orientation with $\beta_G = 0$. By Lemma 2.3, G has a nowhere-zero 3-flow, which is a contradiction. □

4 Proof of Theorem 1.18

Assume not. Suppose that G is a counterexample, such that $|V(G)| + |E(G)|$ is as small as possible. Let $S' = \{x \in V(G) : d_G(x) = 5\}$ and $S = \{C = \partial_G(X) : |C| = 5,\ X \subset V(G)\}$. Let β_G be a Z_3-boundary, such that G has no β_G-orientation.

Claim 4.1. *$|V(G)| \geq 3$ and $|S'| \leq |S| \leq 5$.*

Proof. Since G is 5-edge-connected, $|V(G)| \geq 2$. If $|V(G)| = 2$, let $V(G) = \{x, y\}$, then all the edges of G are between x and y, and $|E(G)| \geq 5$. Let D_x be an orientation of x, such that $d^+_{D_x}(x) - d^-_{D_x}(x) \equiv \beta_G(x) \pmod 3$. Since β_G is a Z_3-boundary, $d^+_{D_x}(y) - d^-_{D_x}(y) \equiv \beta_G(y) \pmod 3$. Therefore G has a β_G-orientation, a contradiction. Hence $|V(G)| \geq 3$ and $|S'| \leq |S| \leq 5$. □

Claim 4.2. *Let $U \subset V(G)$ with $|U| \geq 2$. If $d_G(U) = 5$, then $U \cap S' \neq \emptyset$.*

Proof. If not, choose U to be a minimal one such that: for any $A \subset U$ with $2 \leq |A| < |U|$, we have $d_G(A) \neq 5$.

By the minimality of G, G/U has a β_G-orientation D' which is a partial β_G-orientation of G, such that $d^+_{D'}(x) - d^-_{D'}(x) \equiv \beta_G(x) \pmod 3$ for each $x \in V(G) \setminus U$.

Let G' be a graph obtained from G by contracting $V(G) \setminus U$ as z_0, and let $\beta_{G'} = \beta_G$.

(i) Since $V(G') = U + z_0$, $|V(G')| = |U| + 1 \geq 3$.

(ii) Since $d_{G'}(z_0) = d_G(U) = 5$, by Proposition 2.2 (1), we have that $d_{G'}(z_0) \leq 4 + |\tau_{G'}(z_0)|$.

(iii) By the assumption and minimality of U, we have that for any $A \subset U$, $d_G(A) \neq 5$. Therefore $d_G(A) \geq 6$. By Proposition 2.2 (2), $d_{G'}(A) = d_G(A) \geq 4 + |\tau_{G'}(A)|$.

By Lemma 2.4, the pre-orientation of $E'(z_0)$ of all edges incident with z_0 can be extended to a $\beta_{G'}$-orientation of G'. Therefore, G has a β_G-orientation, which is a contradiction. □

Let G_1' be a graph obtained from G by adding a new vertex z_0 and $|S'|$ arcs from z_0 to S', such that each vertex in S' has degree 6 in G_1'.

Define $\beta_{G_1'}$ as follows:

(1) $\beta_{G_1'}(x) = \beta_G(x)$ if $x \notin S' + z_0$;

(2) $\beta_{G_1'}(x) \equiv \beta_G(x) - 1 \pmod 3$ if $x \in S'$;

(3) $\beta_{G_1'}(z_0) \equiv |S'| \pmod 3$ and $\beta_{G_1'}(z_0) \in \{0, 1, 2\}$.

Now $d_{G_1'}(z_0) \leq 4 + |\tau_{G_1'}(z_0)|$ and $|V(G_1')| = |V(G)| + 1 \geq 4$. We claim that $d_{G_1'}(A) \geq 4 + |\tau_{G_1'}(A)|$, for each nonempty vertex subset A not containing z_0 with $|V(G_1') \setminus A| > 1$.

If $A \cap S' = \emptyset$, then by Claim 4.2, $d_{G_1'}(A) = d_G(A) \neq 5$. Thus $d_{G_1'}(A) \geq 6$. By Proposition 2.2 (2), $d_{G_1'}(A) \geq 4 + |\tau_{G_1'}(A)|$.

If $A \cap S' \neq \emptyset$, then $d_{G_1'}(A) \geq d_G(A) + 1 \geq 6$. By Proposition 2.2 (2), we have that $d_{G_1'}(A) \geq 4 + |\tau_{G_1'}(A)|$.

Now G_1' satisfies all the conditions of Lemma 2.4. By Lemma 2.4, G_1' has a $\beta_{G_1'}$-orientation extended from the pre-orientation of $E_1'(z_0)$ of all edges incident with z_0, which implies that G has a β_G-orientation, a contradiction.

The proof is complete. □

Acknowledgement

The second author is supported by NSFC (No. 11271300) and the Doctorate Foundation of Northwestern Polytechnical University (cx201326).

References

[1] N. Alon, N. Linial and R. Meshulam, Additive bases of vector spaces over prime fields, *J. Combin. Theory Ser. A* **57** (1991) 203–210.

[2] F. Jaeger, Flows and generalized coloring theorems in graphs, *J. Combin. Theory Ser. B* **26** (1979) 205–216.

[3] F. Jaeger, N. Linial, C. Payan and M. Tarsi, Group connectivity of graphs–a nonhomogeneous analogue of nowhere-zero flow properties, *J. Combin. Theory Ser. B* **56** (1992) 165–182.

[4] M. Kochol, An equivalent version of the 3-flow conjecture, *J. Combin. Theory Ser. B* **83** (2001) 258–261.

[5] M. Kochol, Superposition and constructions of graphs without nowhere-zero k-flows, *Europ. J. Combin.* **23** (2002) 281–306.

[6] H.-J. Lai and C.-Q. Zhang, Nowhere-zero 3-flows of highly connected graphs, *Discrete Math.* **110** (1992) 179–183.

[7] L. M. Lovász, C. Thomassen, Y.-Z Wu and C.-Q. Zhang, Nowhere-zero 3-flows and modulo k-orientations, *J. Combin. Theory Ser. B* **103** (2013) 587–598.

[8] C. Thomassen, The weak 3-flow conjecture, *J. Combin. Theory Ser. B* **102** (2012) 521–529.

[9] W. T. Tutte, On the imbedding of linear graphs in surfaces, *Proc. London Math. Soc.* **51** (1949) 474–483.

[10] W. T. Tutte, A contribution on the theory of chromatic polynomial, *Canad. J. Math.* **6** (1954) 80–91.

[11] C.-Q. Zhang, Integer Flows and Cycle Covers of Graphs, *Marcel Dekker Inc.*, New York, (1997).

ARS MATHEMATICA CONTEMPORANEA

Also available at http://amc-journal.eu
ISSN 1855-3966 (printed edn.), ISSN 1855-3974 (electronic edn.)
ARS MATHEMATICA CONTEMPORANEA 10 (2016) 99–112

Quasi-configurations: Building blocks for point – line configurations

Jürgen Bokowski *
Technische Universität Darmstadt, Dolivostraße 15, 64293 Darmstadt, Germany

Vincent Pilaud †
CNRS & LIX, École Polytechnique, 91128 Palaiseau Cedex, France

Received 31 March 2014, accepted 19 March 2015, published online 24 July 2015

Abstract

We study point – line incidence structures and their properties in the projective plane. Our motivation is the problem of the existence of (n_4) configurations, still open for few remaining values of n. Our approach is based on quasi-configurations: point – line incidence structures where each point is incident to at least 3 lines and each line is incident to at least 3 points. We investigate the existence problem for these quasi-configurations, with a particular attention to $3|4$-configurations where each element is 3- or 4-valent. We use these quasi-configurations to construct the first (37_4) and (43_4) configurations. The existence problem of finding (22_4), (23_4), and (26_4) configurations remains open.

Keywords: Projective arrangements, point – line incidence structure, (n_k) configurations.

Math. Subj. Class.: 52C30

1 Introduction

A *geometric (n_k) configuration* is a collection of n points and n lines in the projective plane such that each point lies on k lines and each line contains k points. We recommend our reader to consult Grünbaum's book [7] for a comprehensive presentation and an historical perspective on these configurations. The central problem studied in this book is to determine for a given k those numbers n for which there exist geometric (n_k) configurations. The answer is completely known for $k = 3$ (geometric (n_3) configurations exist if

*http://wwwopt.mathematik.tu-darmstadt.de/~bokowski/

†Corresponding author. http://www.lix.polytechnique.fr/~pilaud/. VP was partially supported by the Spanish MICINN grant MTM2011-22792 and by the French ANR grant EGOS 12 JS02 002 01.

E-mail addresses: juergen.bokowski@gmail.com (Jürgen Bokowski), vincent.pilaud@lix.polytechnique.fr (Vincent Pilaud)

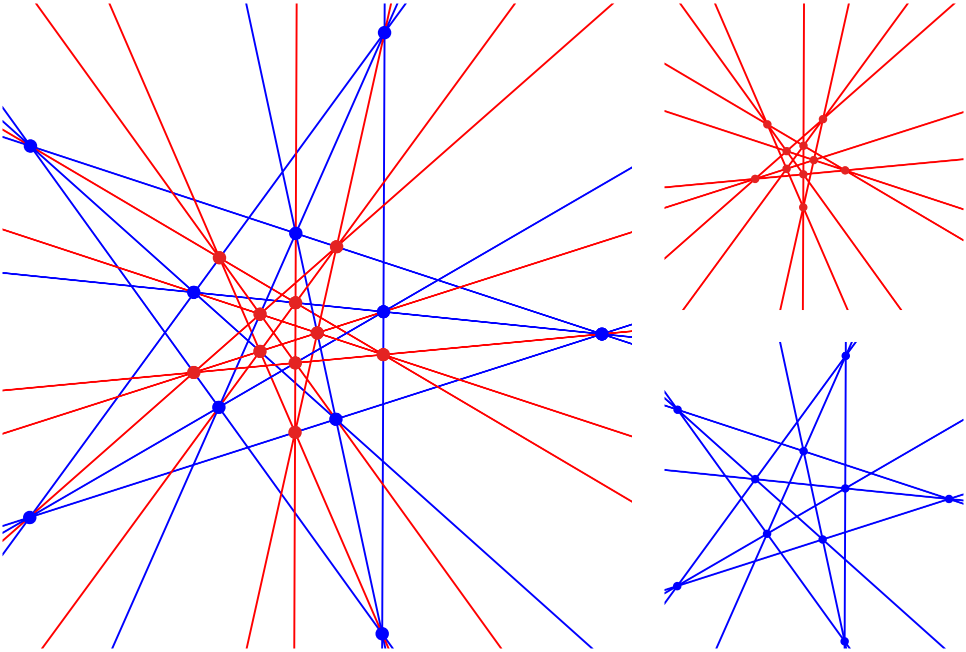


Figure 1: Splitting Grünbaum's geometric (20_4) configuration [6] into two (10_3) configurations.

and only if $n \geq 9$), partially solved for $n = 4$ (geometric (n_4) configurations exist if and only if $n = 18$ or $n \geq 20$ with a finite list of possible exceptions), and wide open for $k > 4$. Our contribution concerns $k = 4$, where we provide solutions for two former open cases: there exist geometric (37_4) and (43_4) configurations. Moreover, we study building blocks for constructing geometric (n_k) configurations that might be of some help for clarifying the final open cases (22_4), (23_4), and (26_4). Many aspects of our presentation appeared during our investigation of the case (19_4) in which there is no geometric (19_4) configuration, see [1, 2].

The approach of this paper is to construct geometric (n_4) configurations from smaller building blocks. For example, Grünbaum's geometric (20_4) configuration [6] can be constructed by superposition of two geometric (10_3) configurations as illustrated in Figure 1. To extend this kind of construction, we study an extended notion of point – line configurations, where incidences are not regular but still prescribed.

1.1 Point – line incidence structures

We define a *point – line incidence structure* as a set P of *points* and a set L of *lines* together with a point – line *incidence relation*, where two points of P can be incident with at most one line of L and two lines of L can be incident with at most one point of P. Throughout the paper, we only consider *connected* incidence structures, where any two elements of $P \sqcup L$ are connected via a path of incident elements.

For a point – line incident structure (P, L), we denote by p_i the number of points of P contained in i lines of L and similarly by ℓ_j the number of lines of L containing j points of P. We find it convenient to encode these incidence numbers into a pair of polynomials $(\mathbf{P}(x), \mathbf{L}(y))$, called the *signature* of (P, L), and defined by

$$\mathbf{P}(x) := \sum\nolimits_i p_i x^i \qquad \text{and} \qquad \mathbf{L}(y) := \sum\nolimits_j \ell_j y^j.$$

For example, the point – line incidence structure represented in Figure 2 has signature $(8x^3 + 2x^4, 8y^3 + 2y^4)$. With these notations, the number of points and lines are given by $|P| = \mathbf{P}(1)$ and $|L| = \mathbf{L}(1)$, while the number of point – line incidences is given by $|\{(p, \ell) \in P \times L \mid p \in \ell\}| = \mathbf{P}'(1) = \mathbf{L}'(1)$.

We distinguish three different levels of point – line incidence structures, in increasing generality:

Geometric Points and lines are ordinary points and lines in the real projective plane $\mathbb{P}$.

Topological Points are ordinary points in $\mathbb{P}$, but lines are *pseudolines*, *i.e.*, non-separating simple closed curves of $\mathbb{P}$ which cross pairwise precisely once.

Combinatorial Just an abstract incidence structure (P, L) as described above, with no additional geometric structure.

In this paper, we are mainly interested in the geometric level. We therefore omit the word geometric in what follows unless we have to distinguish different levels.

1.2 (n_k) configurations

One of the main problems in the theory of point – line incidence structures is to clarify the existence of regular point – line incidence structures. A *k-configuration* is a point – line incidence structure (P, L) where each point of P is contained in k lines of L and each line of L contains k points of P. In such a configuration, the number of points equals the number of lines, and thus it has signature (nx^k, ny^k). If we want to specify the number of points and lines, we call it an (n_k) *configuration*. We refer to the recent monographs of Grünbaum [7] and Pisanski and Servatius [8] for comprehensive presentations of these objects. Classical examples of regular configurations are Pappus' and Desargues' configurations, which are respectively (9_3) and (10_3) configurations. In the study of the existence of (n_4) configurations there are still a few open cases. Namely, it is known that (geometric) (n_4) configurations exists if and only if $n = 18$ or $n \geq 20$, with the possible exceptions of $n = 22, 23, 26, 37$ and 43 [5, 4, 2]. Different methods have been used to obtain the current results on the existence of 4-configurations:

(i) For $n \leq 16$, Bokowski and Schewe [3] used a counting argument based on Euler's formula to prove that there exist no (n_4) configuration, even topological.

(ii) For small values of n, one can search for all possible (n_4) configurations. For $n = 17$ or 18, one can first enumerate all combinatorial (n_4) configurations and search for geometric realizations among them. This approach was used by Bokowski and Schewe in [4] to show that there is no (17_4) configuration and to produce a first (18_4) configuration. Another approach, proposed in [1], is to enumerate directly all topological (n_4) configurations, and to search for geometric realizations among this restricted family. In this way, we showed that there are precisely two (18_4) configurations, that of [4] and another one [1], see Figure 3. For $n = 19$, we obtained in [1] all 4 028 topological (19_4) configurations and the study of their realizability has led to the result that there is no geometric (19_4) configuration [2].

(iii) For larger values of n, one cannot expect a complete classification of (n_4) configurations. However, one can construct families of examples of 4-configurations. One of the key ingredients for such constructions is the use of symmetries. See Figure 1 for the smallest example obtained in this way, and refer to the detailed presentation in Grünbaum's recent monograph [7].

(iv) Finally, Bokowski and Schewe introduced in [4] a method to produce (n_4) configurations from deficient configurations. It consists in finding two point – line incidence structures (P, L) and (P', L') of respective signatures $(ax^3 + bx^4, cy^3 + dy^4)$ and $(cx^3 + ex^4, ay^3 + fy^4)$, where $a + b + c + e = a + c + d + f = n$, and a projective transformation which sends the 3-valent points of P to points contained in a 3-valent line of L', and at the same time the 3-valent lines of L to lines containing a 3-valent point of P'. This method was used to obtain the first examples of (29_4) and (31_4) configurations.

In this paper, we are interested in this very last method described above. We are going to study deficient configurations (see the notion of quasi-configuration and 3|4-configuration in the next subsection) for the use of them as building blocks for configurations. Our study has led in particular to first examples of (37_4) and (43_4) configurations. Thus the remaining undecided cases for the existence of (n_4) configurations are now only the cases $n = 22, 23,$ and 26.

1.3 Quasi-configurations

A *quasi-configuration* (P, L) is a point – line incidence structure in which each point is contained in at least 3 lines and each line contains at least 3 points of P. In other words, the signature $(\mathbf{P}, \mathbf{L})$ of (P, L) satisfies $x^3 \mid \mathbf{P}(x)$ and $y^3 \mid \mathbf{L}(y)$. The term "quasi-configuration" for this concept was suggested by Grünbaum to the authors. As observed above, these quasi-configurations can sometimes be used as building blocks for larger point – line incidence structures.

In this paper, we investigate in particular 3|4-*configurations*, where each point of P is contained in 3 or 4 lines of L and each line of L contains 3 or 4 points of P. In other words, 3|4-configurations are point – line incidence structures whose signature is of the form $(ax^3 + bx^4, cx^3 + dx^4)$ for some $a, b, c, d \in \mathbb{N}$ satisfying $3a + 4b = 3c + 4d$. Note that their numbers of points and lines do not necessarily coincide. If it is the case, *i.e.*, if $a + b = c + d = n$, we speak of an $(n_{3|4})$ *configuration*. In this case, $a = c$ and $b = d$, the number of points and lines is $n = a + b = c + d$ and the number of incidences is $3a + 4b = 3c + 4d$.

We think of an $(n_{3|4})$ configuration as a deficient (n_4) configuration. A good measure on $(n_{3|4})$ configurations is the number of missing incidences a. We say that an $(n_{3|4})$ configuration is *optimal* if it contains the maximal number of point – line incidences among all $(n_{3|4})$ configurations. One objective is to study and classify optimal $(n_{3|4})$ configurations for small values of n.

Example 1.1. Figure 2 shows an incidence structure with signature $(8x^3+2x^4, 8y^3+2y^4)$. It is a $10_{3|4}$-configuration: the 3-valent elements are colored red while the 4-valent elements are colored blue. We will see in Figure 7 that this 3|4-configuration is not optimal.

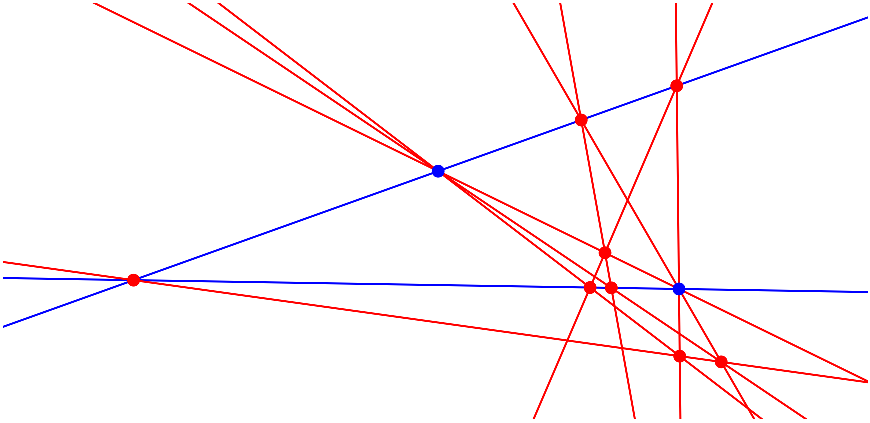


Figure 2: A quasi-configuration with signature $(8x^3 + 2x^4, 8y^3 + 2y^4)$.

1.4 Overview

The paper is divided into two parts. In Section 2, we illustrate how quasi-configurations (in particular $3|4$-configurations) can be used as building blocks to construct (n_4) configurations, and we obtain in particular examples of (37_4) and (43_4) configurations. In Section 3, we present a counting obstruction for the existence of topological quasi-configurations, and we study optimal $(n_{3|4})$ configurations with few points and lines.

2 Constructions

We discuss here different ways to obtain new point – line incidence structures from old ones. We are in particular interested in the construction of new quasi-configurations from old ones. We use these techniques to provide the first (37_4) and (43_4) configurations.

2.1 Operations on point – line incidence structures

To construct new point – line incidence structures from old ones, we will use the following operations, illustrated in Section 2.2:

Deletion Deleting elements from a point – line incidence structure yields a smaller incidence structure. Note that deletions do not necessarily preserve connectedness or quasi-configurations. We can however use deletions in 4-configurations to construct $3|4$-configurations if no remaining element is incident to two deleted elements.

Addition As illustrated by the example of Grünbaum's (20_4) configuration [6] in Figure 1, certain point – line incidence structures can be obtained as the disjoint union of two smaller incidence structures (P, L) and (P', L'). In particular, we obtain an (n_4) configuration if (P, L) and (P', L') are $3|4$-configurations, if each 3-valent element of (P, L) is incident to precisely one 3-valent element of (P', L') and conversely, and if no other incidences appear.

Splitting The reverse operation of addition is splitting: given a point – line incidence structure, we can split it into two smaller incidence structures. We can require additionally

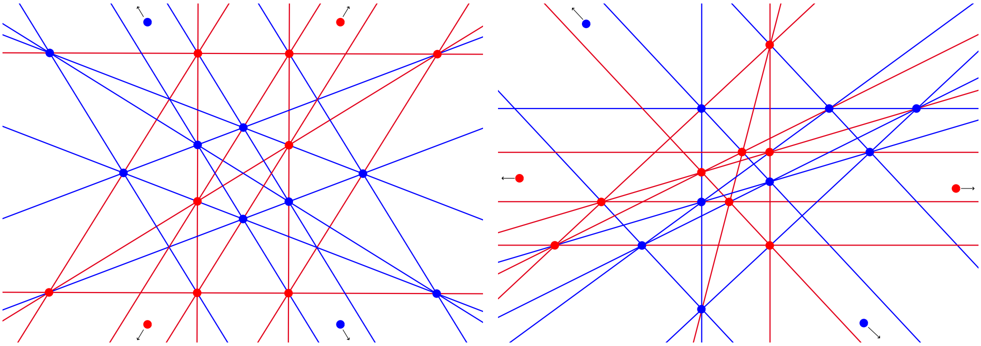


Figure 3: Splittings of the two geometric (18_4) configurations [4, 1] into two $(9_{3|4})$ configurations. The rightmost (18_4) configuration even splits into two (9_3) configurations. The points which seem isolated are in fact at infinity in the direction pointed by the corresponding arrow, and are incident to the 4 lines parallel to that direction.

the two resulting incidence structures to be quasi-configurations or even regular configurations. For example, the two geometric (18_4) configurations [4, 1] as well as Grünbaum's (20_4) configuration [6] are splittable into $3|4$-configurations, see Figures 1 and 3.

Superposition Slightly more general than addition is the superposition, where we allow the two point – line incidence structures (P, L) and (P', L') to share points or lines. For example, we can superpose two 2-valent vertices to make one 4-valent vertex. This idea is used in our construction of (37_4) and (43_4) configurations below.

2.2 Examples of constructions

We now illustrate the previous operations and produce 4-configurations from smaller point – line incidence structures. We start with a simple example.

Example 2.1 (A (38_4) configuration)**.** It was shown in [2] that no topological (19_4) configuration can be geometrically realized with points and lines in the projective plane. However, Figure 4 (left) shows a geometric realization of a topological (19_4) configuration where one line has been replaced by a circle. Forgetting this circle, we obtain a $3|4$-configuration with signature $(15x^4 + 4x^3, 18x^4)$. We take two opposite copies of this $3|4$-configuration (colored purple and red in Figure 4 (right)) and add two lines (colored green in Figure 4 (right)) each incident to two points in each copy. We obtain a (38_4) configuration.

Using similar ideas, we observe that it is always possible to produce a 4-configuration from any $3|4$-configuration.

Example 2.2 (Any $3|4$-configuration generates a 4-configuration)**.** From a $3|4$-configuration with signature $(ax^3 + bx^4, cy^3 + dy^4)$, we can construct an (n_4) configuration where $n = 16a + 16b + 4c = 4a + 16c + 16d$ as follows:

(i) We take four translated copies of the $3|4$-configuration and add suitable parallel lines through all 3-valent points.

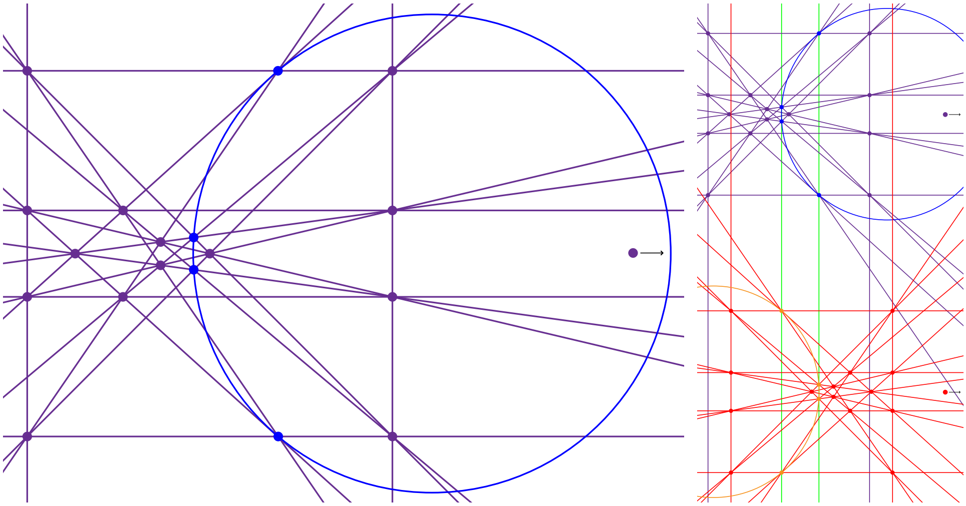


Figure 4: (Left) A geometric realization of a topological (19_4) configuration where one line has been replaced by a circle. (Right) A (38_4) configuration built from two copies of this incidence structure. The construction is explained in full detail in Example 2.1.

(ii) We take the geometric dual of the resulting 3|4-configuration (remember that geometric duality transforms a point p of the projective plane into the line formed by all points orthogonal to p and conversely).

(iii) We take again four translated copies of this dual 3|4-configuration and add suitable parallel lines through all 3-valent vertices.

Of course, we can try to obtain other 4-configurations from 3|4-configurations. This approach was used by Bokowski and Schewe [4] to construct (29_4) and (31_4) configurations from the $(14_{3|4})$, $(15_{3|4})$ and $(16_{3|4})$ configurations of Figure 9. We refer to their paper [4] for an explanation. Here, we elaborate on the same idea to construct two new relevant (n_4) configurations.

Example 2.3 (First (43_4) configuration)**.** To construct a $((n+m)_4)$ configuration from an (n_4) configuration and an (m_4) configuration, we proceed as follows (see Figure 5):

(i) We delete two points not connected by a line in the (n_4) configuration and consider the eight resulting 3-valent lines (colored blue in Figure 5 (top left) and orange in Figure 5 (top right)).

(ii) We add four points (colored green in Figure 5), each incident with precisely two 3-valent lines. All points and lines are now 4-valent again, except the four new 2-valent points.

(iii) We do the same operations in the (m_4) configuration.

(iv) Finally, we use a projective transformation that maps the set of four 2-valent points in the first quasi-configuration onto the set of four 2-valent points in the second quasi-configuration. This transformation superposes the 2-valent points to make them 4-valent.

If this transformation does not superpose other elements than the 2-valent ones and does not create additional unwanted incidences, it yields the desired $((n+m)_4)$ configuration. This construction is illustrated on Figure 5, where we obtain a (43_4) configuration from a (25_4) configuration [7] and a (18_4) configuration [1].

Unfortunately, the method from the previous example cannot provide a (37_4) configuration since there is no (n_4) configuration for $n \leq 17$ [4] and for $n = 19$ [2]. We therefore need another method, which we describe in the following example.

Example 2.4 (First (37_4) configuration)**.** To construct a $((n+m-1)_4)$ configuration from an (n_4) configuration and an (m_4) configuration, we proceed as follows (see Figure 6):

(i) We delete two points on the same line (colored green in Figure 6) of the (n_4) configuration and consider the six resulting 3-valent lines (colored blue in Figure 6 (top left) and orange in Figure 6 (top right)).

(ii) We add three points (colored green in Figure 6), each incident with precisely two 3-valent lines. All points and lines are now 4-valent again, except the initial 2-valent line and the three new 2-valent points.

(iii) We do the same operations in the (m_4) configuration.

(iv) Finally, we use a projective transformation that maps the set of four 2-valent elements in the first quasi-configuration onto the set of four 2-valent elements in the second quasi-configuration. This transformation superposes the 2-valent elements to make them 4-valent.

If this transformation does not superpose other elements than the 2-valent ones and does not create additional unwanted incidences, it yields the desired $((n+m-1)_4)$ configuration. This construction is illustrated on Figure 6, where we obtain a (37_4) configuration from a (20_4) configuration [7] and a (18_4) configuration [1].

We invite the reader to try his own constructions, similar to the constructions of Examples 2.3 and 2.4, using the operations on point – line incidence structures described above. In this way, one can obtain many (n_4) configurations for various values of n. Additional features can even be imposed, such as non-trivial motions or symmetries. We have however not been able to find answers to the following question.

Question 1. Can we create a (22_4) configuration by glueing two quasi-configurations with 11 points and lines each? More generally, can we construct (22_4), (23_4), or (26_4) configurations by superposition of smaller quasi-configurations?

3 Obstructions and optimal 3|4-configurations

In this section, we further investigate point – line incidence structures and 3|4-configurations. We start with a necessary condition for the existence of topological incidence structures with a given signature. For this, we extend to all topological incidence structures an argument of Bokowski and Schewe [3] that was used to prove the non-existence of (15_4) configurations. We obtain the following inequality.

Proposition 3.1. *Let* (P, L) *be topological incidence structure with signature* $(\mathbf{P}, \mathbf{L})$. *Then*

$$\mathbf{P}''(1) + 2\mathbf{P}'(1) - \mathbf{L}(1)^2 + \mathbf{L}(1) - 6\mathbf{P}(1) + 6 \leq 0.$$

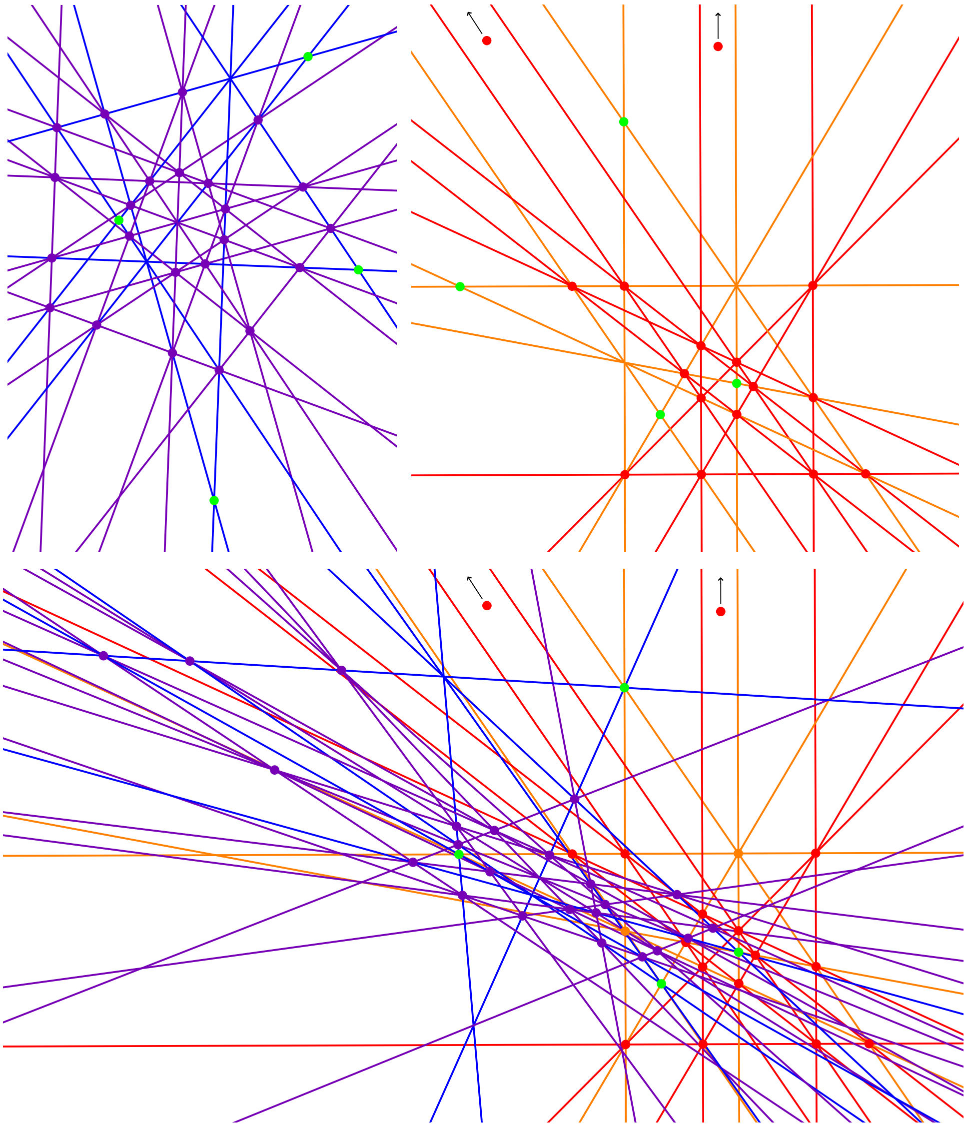


Figure 5: A (43_4) configuration built from deficient (25_4) and (18_4) configurations. The construction is explained in full detail in Example 2.3.

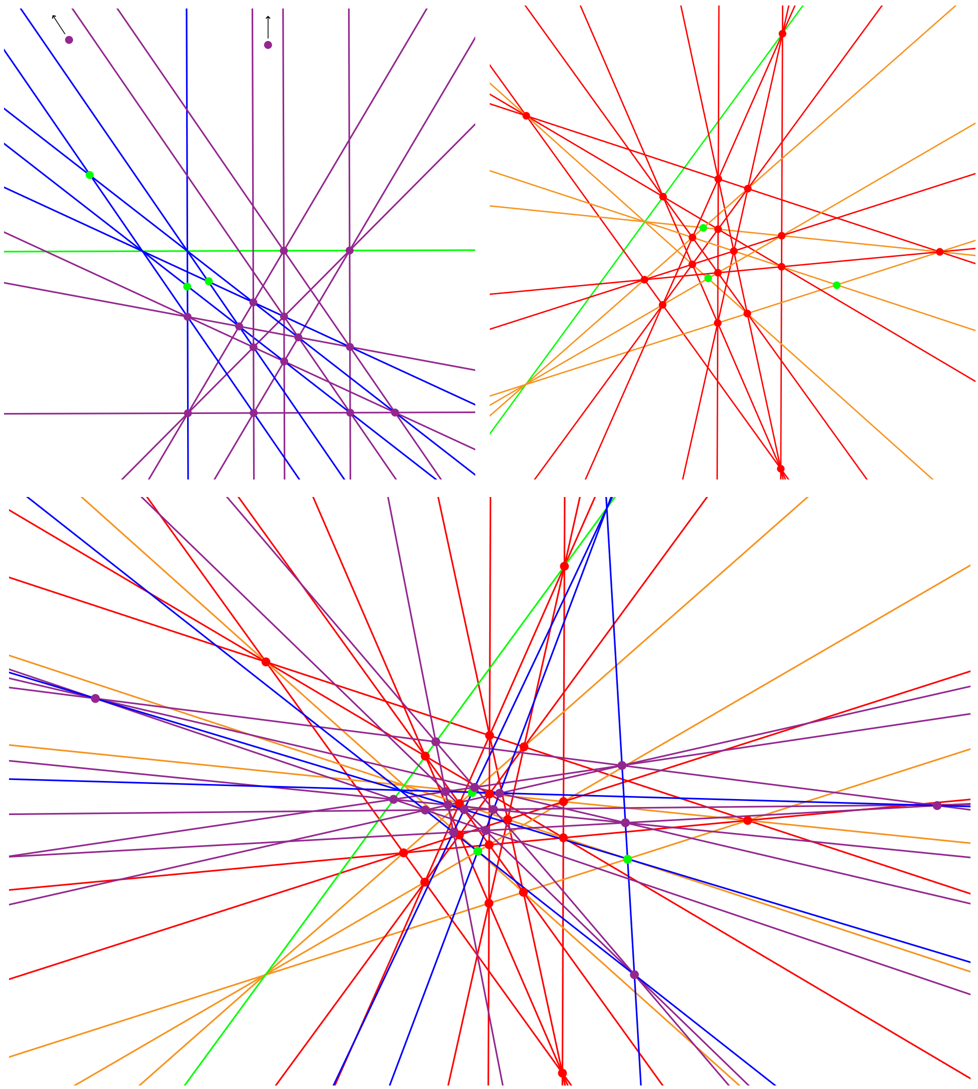


Figure 6: A (37_4) configuration built from deficient (20_4) and (18_4) configurations. The construction is explained in full detail in Example 2.4.

Proof. Let p_i be the number of i-valent points and ℓ_j the number of j-valent lines in the incidence structure (P, L). The signature $(\mathbf{P}, \mathbf{L})$ is $\mathbf{P}(x) := \sum_i p_i x^i$ and $\mathbf{L}(y) := \sum_j \ell_j y^j$.

Since the incidence structure is topological, we can draw it on the projective plane such that no three pseudolines pass through a point which is not in P. We call *additional* 2-*crossings* the intersection points of two lines of L which are not points of P. We consider the lifting of this drawing on the 2-sphere. We obtain a graph embedded on the sphere, whose vertices are all points of P together with all additional 2-crossings, whose edges are the segments of lines of L located between two vertices, and whose faces are the connected components of the complement of L. Let f_0, f_1 and f_2 denote respectively the number of vertices, edges and faces of this map. Denoting by $\deg(p)$ the number of lines of L containing a point $p \in P$ and similarly by $\deg(\ell)$ the numbers of points of P contained in a line $\ell \in L$, we have

$$f_0 = 2\binom{\mathbf{L}(1)}{2} - 2\sum_{p\in P}\left(\binom{\deg(p)}{2} - 1\right) = \mathbf{L}(1)\big(\mathbf{L}(1)-1\big) + 2\mathbf{P}(1) - \sum_i i(i-1)p_i,$$

$$f_1 = 2\sum_{\ell\in L}\deg(\ell) + 2f_0 - 2\mathbf{P}(1) = 2\sum_j j\ell_j + 2f_0 - 2\mathbf{P}(1) = 2\sum_i ip_i + 2f_0 - 2\mathbf{P}(1),$$

$$f_2 = f_1 - f_0 + 2.$$

Moreover, since no face is a digon, we have $3f_2 \leq 2f_1$. Replacing f_2 and f_1 by the above expressions, we obtain

$$\begin{aligned} 0 \geq 3f_2 - 2f_1 = f_1 - 3f_0 + 6 &= 2\sum_i ip_i - 4\mathbf{P}(1) - f_0 + 6 \\ &= \sum_i i(i+1)p_i - \mathbf{L}(1)\big(\mathbf{L}(1)-1\big) - 6\big(\mathbf{P}(1)-1\big), \end{aligned}$$

and thus the desired inequality. □

Corollary 3.2. *If $(ax^3 + bx^4, ay^3 + by^4)$ is the signature of a topological incidence structure, then*

$$-(a+b)^2 + 7a + 15b + 6 \leq 0.$$

The following table provides the minimum value of b for which there could exist a topological incidence structure with signature $(ax^3 + bx^4, ay^3 + by^4)$:

a	0	1	2	3	4	5	6	7
$b_{\min}$	16	14	13	11	9	8	6	3

Proof. Direct from Proposition 3.1 with $\mathbf{P}(x) = ax^3 + bx^4$ and $\mathbf{L}(y) = ay^3 + by^4$. □

For example, there is no topological (15_4) configuration [3] and no incidence structure with signature $(7x^3 + 2x^4, 7y^3 + 2y^4)$. Compare to Example 1.1 which shows that a configuration with signature $(8x^3 + 2x^4, 8y^3 + 2y^4)$ exists.

Corollary 3.3. *A $(n_{3|4})$ configuration has at most* $I_{\max} := \min\left(4n\ ,\ \left\lfloor \frac{n^2+17n-6}{8} \right\rfloor\right)$ *incidences. The values of $I_{\max}$ appear in the following table:*

n	7	8	9	10	11	12	13	14	15	16
$I_{\max}$	20	24	28	33	37	42	48	53	59	64

Proof. Consider an $(n_{3|4})$ configuration with signature $(ax^3 + bx^4, ay^3 + by^4)$ where $a + b = n$. The number of incidences is $I := 3a + 4b$. It can clearly not exceed $4n$. For the second term in the minimum, we apply Corollary 3.2 to get

$$\begin{aligned} 0 &\geq -(a+b)^2 + 7a + 15b + 6 \\ &= -(a+b)^2 + 8(3a+4b) - 17(a+b) + 6 \\ &= -n^2 + 8I - 17n + 6. \end{aligned}$$

□

Corollary 3.4. *There is no topological $(n_{3|4})$ configuration if $n \leq 8$.*

Proof. If $n \leq 7$, there is no topological $(n_{3|4})$ configuration since it should have at least $3n$ incidences, which is larger than the upper bound of Corollary 3.3. If $n = 8$, a $(8_{3|4})$ configuration should be a (8_3) configuration by Corollary 3.3. But the only combinatorial (8_3) configuration is not topological. □

To close this section, we exhibit optimal $(n_{3|4})$ configurations for small values of n, *i.e.*, $(n_{3|4})$ configurations which maximize the number of point – line incidences.

Proposition 3.5. *For $9 \leq n \leq 13$, the bound of Corollary 3.3 is tight: there exists $(n_{3|4})$ configurations with $\left\lfloor \frac{n^2+17n-6}{8} \right\rfloor$ incidences.*

Proof. For $n = 13$, we consider the $(13_{3|4})$-configuration of Figure 7. The homogeneous coordinates of its points and lines are given by

$$P := L := \left\{ \begin{bmatrix} i \\ j \\ 1 \end{bmatrix} \;\middle|\; i, j \in \{-1, 0, 1\} \right\} \cup \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \right\}.$$

For $n = 10, 11$ or 12, we obtain $(n_{3|4})$ configurations by removing suitable points and lines in our $(13_{3|4})$ configuration. The resulting configurations are illustrated in Figure 7. (Note that for $n = 10$, we even have two dual ways to suitably remove three points and three lines from our $(13_{3|4})$ configuration: either we remove three 3-valent points and the three 4-valent lines containing two of these points, or we remove three 3-valent lines and the three 4-valent points contained in two of these lines). Finally, for $n = 9$ we use the bottom rightmost $(9_{3|4})$ configuration of Figure 7. □

As a curiosity, we give another example of an optimal $(12_{3|4})$ configuration which contains Pappus' and Desargues' configurations simultaneously. See Figure 8.

Observe that optimal $(n_{3|4})$ configurations are given by (n_4) configurations for large n, and that the only remaining cases for optimal $(n_{3|4})$ configurations are for $n = 14$, 15, 16, 17, 19, 22, 23, and 26. We have represented in Figure 9 some $(15_{3|4})$ and $(16_{3|4})$ configurations which we expect to be optimal, although they do not reach the theoretical upper bound of Corollary 3.3. Observe also that deleting the circle in Figure 4 (left) and adding one line through two of the resulting 3-valent points provides a $(19_{3|4})$ configuration with 74 incidences, which is almost optimal since there is no (19_4) configuration [1, 2]. To conclude, we thus leave the following question open.

Question 2. What are the optimal $(14_{3|4})$ configurations? Are the $(15_{3|4})$ and $(16_{3|4})$ configurations in Figure 9 optimal? Is there a $(19_{3|4})$ configuration with 75 incidences.

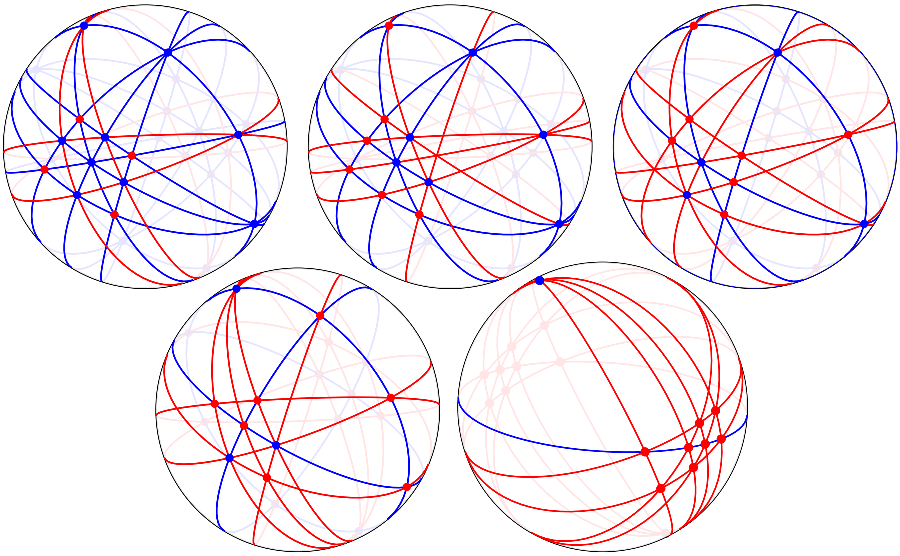


Figure 7: Optimal $(n_{3|4})$ configurations, for $n = 13, 12, 11, 10, 9$. They have respectively 48, 42, 37, 33, and 28 point – line incidences. The 3-valent elements are colored red while the 4-valent elements are colored blue.

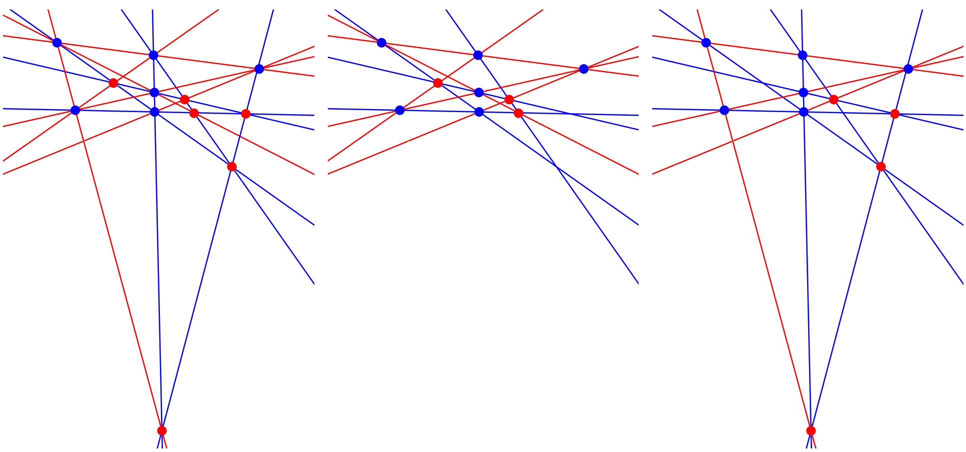


Figure 8: An optimal $(12_{3|4})$ configuration (left) which contains simultaneously Pappus' configuration (middle) and Desargues' configuration (right). In the $(12_{3|4})$ configuration, the 3-valent elements are colored red while the 4-valent elements are colored blue. In the Pappus' and Desargues' subconfigurations, all elements are 3-valent, but we keep the color to see the correspondence better.

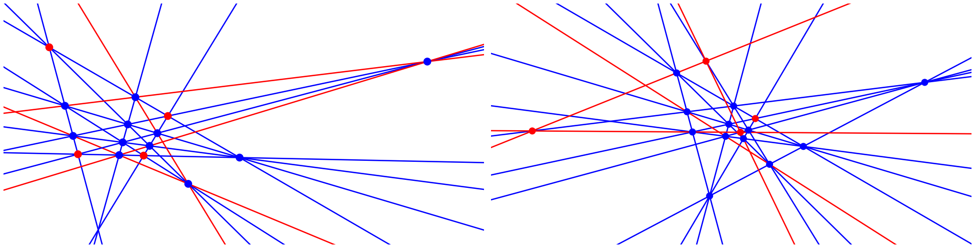


Figure 9: Apparently optimal $(15_{3|4})$ and $(16_{3|4})$ configurations. They have 56 and 60 point – line incidences respectively. The 3-valent elements are colored red while the 4-valent elements are colored blue.

References

[1] J. Bokowski and V. Pilaud, Enumerating topological (n_k)-configurations, *Comput. Geom.* **47** (2014), 175–186.

[2] J. Bokowski and V. Pilaud, On topological and geometric (19_4) configurations, *European J. Combin.* **50** (2015), 4–17.

[3] J. Bokowski and L. Schewe, There are no realizable 15_4- and 16_4-configurations, *Rev. Roumaine Math. Pures Appl.* **50** (2005), 483–493.

[4] J. Bokowski and L. Schewe, On the finite set of missing geometric configurations (n_4), *Comput. Geom.* **46** (2013), 532–540.

[5] B. Grünbaum, Connected (n_4) configurations exist for almost all n, *Geombinatorics* **10** (2000), 24–29.

[6] B. Grünbaum, Musings on an example of Danzer's, *European J. Combin.* **29** (2008), 1910–1918.

[7] B. Grünbaum, *Configurations of points and lines*, volume 103 of *Graduate Studies in Mathematics*, American Mathematical Society, Providence, RI, 2009.

[8] T. Pisanski and B. Servatius, *Configurations from a graphical viewpoint*, Birkhäuser Advanced Texts: Basler Lehrbücher, Birkhäuser/Springer, New York, 2013.

ARS MATHEMATICA CONTEMPORANEA

Also available at http://amc-journal.eu
ISSN 1855-3966 (printed edn.), ISSN 1855-3974 (electronic edn.)
ARS MATHEMATICA CONTEMPORANEA 10 (2016) 113–134

On the split structure of lifted groups

Aleksander Malnič *
Univerza v Ljubljani, Pedagoška fakulteta, Kardeljeva pl. 16, 1000 Ljubljana, Slovenia
Univerza na Primorskem, Inštitut Andrej Marušič, Muzejski trg 2, 6000 Koper, Slovenia

Rok Požar †
Univerza v Ljubljani, Pedagoška fakulteta, Kardeljeva pl. 16, 1000 Ljubljana, Slovenia
Univerza na Primorskem, FAMNIT, Glagoljaška 8, 6000 Koper, Slovenia

Received 18 May 2014, accepted 23 November 2014, published online 24 July 2015

Abstract

Let $\wp\colon \tilde{X} \to X$ be a regular covering projection of connected graphs with the group of covering transformations $\mathrm{CT}_\wp$ being abelian. Assuming that a group of automorphisms $G \le \mathrm{Aut}\, X$ lifts along $\wp$ to a group $\tilde{G} \le \mathrm{Aut}\, \tilde{X}$, the problem whether the corresponding exact sequence $\mathrm{id} \to \mathrm{CT}_\wp \to \tilde{G} \to G \to \mathrm{id}$ splits is analyzed in detail in terms of a Cayley voltage assignment that reconstructs the projection up to equivalence.

In the above combinatorial setting the extension is given only implicitly: neither $\tilde{G}$ nor the action $G \to \mathrm{Aut}\,\mathrm{CT}_\wp$ nor a 2-cocycle $G \times G \to \mathrm{CT}_\wp$, are given. Explicitly constructing the cover $\tilde{X}$ together with $\mathrm{CT}_\wp$ and $\tilde{G}$ as permutation groups on $\tilde{X}$ is time and space consuming whenever $\mathrm{CT}_\wp$ is large; thus, using the implemented algorithms (for instance, `HasComplement` in MAGMA) is far from optimal. Instead, we show that the minimal required information about the action and the 2-cocycle can be effectively decoded directly from voltages (without explicitly constructing the cover and the lifted group); one could then use the standard method by reducing the problem to solving a linear system of equations over the integers. However, along these lines we here take a slightly different approach which even does not require any knowledge of cohomology. Time and space complexity are formally analyzed whenever $\mathrm{CT}_\wp$ is elementary abelian.

Keywords: Algorithm, abelian cover, Cayley voltages, covering projection, graph, group extension, group presentation, lifting automorphisms, linear systems over the integers, semidirect product.

Math. Subj. Class.: 05C50, 05C85, 05E18, 20B40, 20B25, 20K35, 57M10, 68W05

* Corresponding author. Supported in part by "Agencija za raziskovalno dejavnost Republike Slovenije", research program P1-0285, and research projects J1-5433 and J1-6720.

† Supported in part by "Agencija za raziskovalno dejavnost Republike Slovenije", research program P1-0285. The author also thanks the University of Auckland, New Zealand, for hospitality during his 6 months research visit in 2011-12.

1 Introduction

A large part of algebraic graph theory is devoted to analyzing structural properties of graphs with prescribed degree of symmetry in order to classify, enumerate, construct infinite families, and to produce catalogs of particular classes of interesting graphs up to a certain reasonable size. References are too numerous to be listed here, but see for instance [2, 6, 8, 9, 10, 11, 12, 14, 16, 24, 25, 26, 29, 31, 36, 39, 40, 42, 46, 47, 52, 53, 55], and the references therein.

It is not surprising, then, that the techniques employed in these studies are fairly rich and diverse, ranging from pure combinatorial and computational methods to methods from abstract group theory, permutation groups, combinatorial group theory, linear algebra, representation theory, and algebraic topology.

Covering space techniques, and lifting groups of automorphisms along regular covering projections in particular, play a prominent role in this context. (See Section 2 for exact definitions of all notions used in this Introduction.) The idea goes back to Djoković [9] (and to an unpublished work of Conway, see [3, Corollary 19.6]), who constructed first examples of infinite families of graphs of small valency and maximal degree of transitivity, and to Biggs [3, Proposition 19.3], who gave a sufficient condition for a group of automorphisms to lift as a semidirect product. While Djoković's approach is classical in terms of fundamental groups, Biggs expressed his particular lifting condition combinatorially. A combinatorial approach to covering projections of graphs in terms of voltages was systematically developed in the early 70' by Gross and Tucker, see [20], after having been introduced by Alpert and Gross [18, 19] in the context of maps on surfaces.

A systematic combinatorial treatment of lifting automorphisms along covering projections (either in the context of graphs, maps on surfaces, or cell complexes) has been considered by several authors, see [1, 21, 32, 33, 48, 50] and the references therein. More specific types of covers, say, with cyclic or elementary abelian groups of covering transformations, have been extensively studied in [22, 35, 37, 49]; for the applications we refer the reader to [8, 11, 13, 14, 26, 27, 28, 30, 36, 40, 41, 52, 55]. For some recent results on arc transitive cubic graphs arising as regular covers with an abelian group of covering transformations we refer the reader to [7].

Basic lifting techniques in terms of voltages are now well understood, yet several important issues still remain to be considered. In view of the fact that structural properties of graphs often rely on the structure and the action of their automorphism groups, one such topic is investigating the structure of lifted groups – although certain particular questions along these lines have been addressed, see [3, 15, 33, 51]. Other points of interest are algorithmic and complexity aspects of lifting automorphisms, which have so far received even less attention. Certain aspects, but not those considered here, were touched upon in [34, 48].

Specifically, let $\wp\colon \tilde{X} \to X$ be a regular covering projection of connected graphs given in terms of voltages. Assuming that a group $G \leq \operatorname{Aut} X$ lifts along $\wp$, it is of particular interest to study the corresponding exact sequence $\mathrm{id} \to \mathrm{CT}_\wp \to \tilde{G} \to G \to \mathrm{id}$. A natural question in this context is to ask whether the extension is split: on one hand, split extensions are the most easy ones to analyze, while on the other hand, a restrictive situation stemming from the fact that the group G, acting on X, acts also on $\tilde{X}$ via its isomorphic complement

E-mail addresses: aleksander.malnic@guest.arnes.si (Aleksander Malnič), pozar.rok@gmail.com (Rok Požar)

$\overline{G}$ to $\mathrm{CT}_\wp$ within $\tilde{G}$, implies that a lot more information about symmetry properties of $\tilde{X}$ can be derived; moreover, split extensions are frequently encountered in many concrete examples of graph covers. Describing efficient methods for testing whether a given group lifts as a split extension of $\mathrm{CT}_\wp$ is the main objective of this paper.

Methods for testing whether a given extension $1 \to K \to E \to Q \to 1$ of finite groups is split, are known, see [5] and [23, Chapters 7 and 8]. In some way or another, all these methods use the fact that a set of coset representatives of K in E is a complement to K if and only if these representatives satisfy the defining relations of Q.

The essential case to be resolved in the first place is that of K being (elementary) abelian. The idea is to modify an arbitrarily chosen set of coset representatives of K so that the defining relations of Q are satisfied, if possible. Since K is normal and abelian, this modification can be traced in the frame of a certain group algebra, which finally leads to a system of linear equations over the integers (or rather, over prime fields); the complement exists if and only if such a system has a solution.

In practice, the extension can be given in several different ways: either (i) in terms of an epimorphism $E \to Q$, or (ii) in terms of E and the generators of a normal subgroup K, or (iii) via an action $\theta\colon Q \to \mathrm{Aut}\, K$ together with a 2-cocycle $\tau\colon Q \times Q \to K$. In cases (i) and (ii), an essential requirement is that one must have enough information about the extended group E; at least one must know its generators and must be able to perform basic computations in E. In contrast with (i) and (ii), explicit knowledge about E is not needed in case (iii) since the extension can be, up to equivalence of extensions, reconstructed as $K \times Q$ with multiplication rule $(a, x)(b, y) = (a + \theta_x(b) + \tau(x, y), xy)$.

In our setting of graph covers, however, the situation is different and does not fall in any of the above three cases. Namely, the extension $\mathrm{id} \to \mathrm{CT}_\wp \to \tilde{G} \to G \to \mathrm{id}$ is given only implicitly: all the information is encoded in the base graph in terms of voltages that allow G to lift; in particular, neither $\tilde{G}$ nor the action of G on $\mathrm{CT}_\wp$ nor a 2-cocycle are given. Naively translating our setting into the frame of (i) or (ii) and then applying the algorithm already implemented in MAGMA [4] in terms of permutation groups would mean to first compute the covering graph $\tilde{X}$ together with $\mathrm{CT}_\wp$ and $\tilde{G}$ acting on $\tilde{X}$, which, unfortunately, is time and space consuming whenever $\mathrm{CT}_\wp$ is large.

Our situation best fits into the frame of (iii). But in order to follow the approach described in [5] and [23, Chapters 7 and 8] we first need to compute the action $G \to \mathrm{Aut}\,\mathrm{CT}_\wp$ and the 2-cocycle $G \times G \to \mathrm{CT}_\wp$. As we here show, the minimal required information about these data can indeed be effectively decoded directly from voltages (without explicitly constructing the cover and the lifted group). In the actual algorithm, however, we take an approach which is slightly different and even does not require any knowledge of cohomology. Namely, instead of modifying an initial transversal and working within an appropriate group algebra, a potential complement is constructed directly in terms of certain parameters – in view of the fact that a lift of an automorphism is uniquely determined by the mapping of a single vertex – from which the required system of equations is obtained. Although the method works whenever $\mathrm{CT}_\wp$ is abelian, it can be adapted – similarly as in the general context – to treat the case when $\mathrm{CT}_\wp$ is solvable as well.

The paper is organized as follows. In Section 2 we review some basic facts about regular covering projections and lifting automorphisms. In Section 3 we show how to recapture the lifted group $\tilde{G}$ as a crossed product of $\mathrm{CT}_\wp$ by G via reconstructing the coupling $G \to \mathrm{Out}\ \mathrm{CT}_\wp$ and the factor set $G \times G \to \mathrm{CT}_\wp$ in terms of voltages, see Theorem 3.1. In Section 4 we give the necessary and sufficient conditions for G to lift as a split extension of $\mathrm{CT}_\wp$, see Theorem 4.1, and a similar result regarding direct product extensions, see Theorem 4.3. In Section 5 we provide an algorithm for testing whether the extension is split whenever $\mathrm{CT}_\wp$ is abelian, see Subsection 5.1 and Theorem 5.2 in particular. The special case when $\mathrm{CT}_\wp$ is elementary abelian together with formal analysis of time and space complexity is treated in Subsection 5.2, see Theorem 5.6. In Subsection 5.3 we briefly mention the case when $\mathrm{CT}_\wp$ is solvable. Concluding remarks are given in Section 6. For a more substantial account on complexity issues and further applications we refer the reader to [44, 45].

2 Preliminaries

Graphs. Formally, a graph is an ordered 4-tuple $X = (D, V; \mathrm{beg}, ^{-1})$, where $D(X) = D$ and $V(X) = V$ are disjoint sets of *darts* and *vertices*, respectively, beg is the function assigning to each dart its *initial vertex*, and $^{-1}$ is an arbitrary involution on D that creates *edges* arising as orbits of $^{-1}$. For a dart x, its *terminal vertex* is the vertex $\mathrm{end}(x) = \mathrm{beg}(x^{-1})$. An edge $e = \{x, x^{-1}\}$ is called a *link* whenever $\mathrm{beg}(x) \neq \mathrm{end}(x)$. If $\mathrm{beg}(x) = \mathrm{end}(x)$, then the respective edge is either a *loop* or a *semi-edge*, depending on whether $x \neq x^{-1}$ or $x = x^{-1}$, respectively.

There are several reasons for treating graphs formally in a manner just described. For one thing, it is quite versatile for writing down formal proofs; moreover it is indeed natural, even necessary, to consider graphs with semi-edges in different contexts, for instance when dealing with graph covers or when studying graphs that are embedded into surfaces. For a nice use of semi-edges in the context of Cayley graphs we refer the reader to [17].

A *walk* $W\colon u \to v$ of *length* $n \geq 0$ from a vertex $u_0 = u$ to a vertex $u_n = v$ is a sequence of vertices and darts $W = u_0\, x_1\, u_1\, x_2\, u_2\, \ldots\, u_{n-1}\, x_n\, u_n$ where $\mathrm{beg}(x_j) = u_{j-1}$ and $\mathrm{end}(x_j) = u_j$ for all indices $j = 1, \ldots, n$. Its *inverse walk* $W^{-1}\colon v \to u$ is the walk obtained by listing the vertices and darts appearing in W in reverse order. The walk u is the *trivial walk* at the vertex u. Walks of length 1 are sometimes referred to as *arcs*. A graph is *connected* if any two vertices are connected by a walk. A walk is *reduced* if no two consecutive darts in the walk are inverse to each other. Clearly, each walk W has an associated reduced walk $\underline{W}$ obtained by recursively deleting all appearances $uxvx^{-1}$ of consecutive pairs of inverse darts (together with the respective vertices). Two walks $W, W'\colon u \to v$ with the same reduction are called *homotopic*. *Homotopy* is an equivalence relation on the set of all walks, with homotopy classes denoted by $[W]$. Observe that the naturally defined product of walks W_1W_2 by 'concatenation', when defined, carries over to homotopy classes, $[W_1][W_2] = [W_1W_2]$. Assuming the graph X to be connected, the set of homotopy classes of *closed walks* $u \to u$, equipped with the above product, defines the *first homotopy group* $\pi(X, u)$. The trivial class $1_u = [u]$ consists of all walks *contractible* to u. Note that the isomorphism class of $\pi(X, u)$ does not depend on u. More precisely, $\pi(X, u)$ is isomorphic to the free product of cyclic groups $\mathbb{Z}$ or $\mathbb{Z}_2$ (where the $\mathbb{Z}_2$ factors correspond bijectively to the set of all semi-edges in X). A generating set for $\pi(X, u)$ is provided by *fundamental closed walks at* u relative to an arbitrarily chosen spanning tree.

Let X and X' be graphs. A *graph homomorphism* $f\colon X \to X'$ is an adjacency preserving mapping taking darts to darts and vertices to vertices, or more precisely, $f(\mathrm{beg}(x)) = \mathrm{beg}(f(x))$ and $f(x^{-1}) = f(x)^{-1}$. Homomorphisms are composed as functions, $(fg)(x) = f(g(x))$. Given a graph X we frequently need to consider the restricted (and the induced) left action of its group of automorphisms $\mathrm{Aut}\, X$ on certain subsets of X, for instance the sets of vertices, darts, edges etc. As $\mathrm{Aut}\, X$ is by definition a permutation group on the union $V(X) \cup D(X)$ of disjoint sets of vertices and darts, it acts faithfully on $V(X) \cup D(X)$. However, its action on $V(X)$ need not be faithful unless X is *simple*, that is, if it has no parallel links, loops, or semi-edges. We say that a group $G \leq \mathrm{Aut}\, X$ acts *semiregularly on* X whenever it acts freely on $V(X)$ (meaning that if $g \in G$ fixes a vertex it must be the identity on vertices and darts).

Covers. To fix the notation and terminology, and for easier reading, we quickly review some essential facts about covers. The interested reader is referred to [20, 33, 54] for more information.

A *covering projection* of graphs is a surjective homomorphism $\wp\colon \tilde{X} \to X$ mapping the set of darts with a common initial vertex in the *covering graph* $\tilde{X}$ bijectively to the set of darts at the image of that vertex in the *base graph* X. The preimages $\mathrm{fib}_u = \wp^{-1}(u)$, $u \in V(X)$, and $\mathrm{fib}_x = \wp^{-1}(x)$, $x \in D(X)$, are the *vertex-* and *dart-fibres*, respectively. From the definition of a covering projection it immediately follows that for any walk $W\colon u \to v$ in X and an arbitrary vertex $\tilde{u} \in \mathrm{fib}(u)$ there is a unique *lifted walk* $\tilde{W}^{\tilde{u}}$ with $\mathrm{beg}(\tilde{W}^{\tilde{u}}) = \tilde{u}$ that projects to W. This is known as the *unique-path lifting property*. Consequently, if X is connected (which will be our standard assumption without loss of generality) then all fibres have equal cardinality, usually referred to as the *number of folds*. It is also immediate that homotopic walks lift to homotopic walks, and that $\tilde{u} \cdot [W] = \mathrm{end}(\tilde{W}^{\tilde{u}})$ defines a 'right action' of homotopy classes on the vertex set of $\tilde{X}$. In particular, the fundamental group $\pi(X, u)$ acts on the right on fib_u, with the stabilizer of $\tilde{u} \in \mathrm{fib}_u$ being isomorphic to $\pi(\tilde{X}, \tilde{u})$. It is precisely this action that is responsible for the structural properties of the covering.

Covering projections that are particularly important when studying symmetry properties of covers are the *regular covering projections*. By definition, a covering projection $\wp\colon \tilde{X} \to X$ is *regular* if there exists a semiregular group $C \leq \mathrm{Aut}\, \tilde{X}$ such that its orbits on vertices and on darts coincide with vertex- and dart-fibres, respectively. In other words, C acts *regularly* on each fibre (hence the name), and so the covering projection is $|C|$-fold.

Regular covering projections can be grasped combinatorially as follows. First of all, given a graph X and an (abstract) group Γ, let $\zeta\colon D(X) \to \Gamma$ be a function such that $\zeta(x^{-1}) = (\zeta(x))^{-1}$. (For convenience we shall write $\zeta_x = \zeta(x)$ and $\zeta_x^{-1} = (\zeta_x)^{-1}$.) In this context, Γ is called a *voltage group*, ζ is a *Cayley voltage assignment* on X, and ζ_x is the *voltage* of the dart x. We remark that a voltage assignment as above is known as an *ordinary voltage assignment* in the literature [20]. With these data we may define the *derived graph* $\mathrm{Cov}(\zeta)$ with vertex set $V(X) \times \Gamma$ and dart set $D(X) \times \Gamma$, where $\mathrm{beg}(x, c) = (\mathrm{beg}(x), c)$ and $(x, c)^{-1} = (x^{-1}, c\,\zeta_x)$. The projection onto the first coordinate defines a regular covering projection $\wp_\zeta\colon \mathrm{Cov}(\zeta) \to X$. The required semiregular group C is obtained by viewing $C = \Gamma$ as a group of automorphisms of $\mathrm{Cov}(\zeta)$ via its left action on the second coordinates by left multiplication on itself: an element $a \in \Gamma$ maps the vertex (u, c) to (u, ac) and the dart (x, c) to (x, ac). In addition, call the right action of Γ on itself by right multiplication a *voltage-action*. This action determines how walks of length 1 lift: a walk uxv lifts to walks $(u, c)(x, c)(v, c\,\zeta_x)$, for $c \in \Gamma$.

Conversely, with any regular covering projection $\wp\colon \tilde{X} \to X$ we can associate a Cayley voltage assignment ζ on X such that $\wp_\zeta\colon \mathrm{Cov}(\zeta) \to X$ 'essentially reconstructs' the projection $\wp$ in a sense to be described below. Indeed, let $\Gamma = C$ be the semiregular group from the definition of a regular covering. As Γ acts regularly on each fibre we may label the vertices and the darts of $\tilde{X}$ by elements of $V(X) \times \Gamma$ and $D(X) \times \Gamma$, respectively, as follows. Choosing arbitrarily a vertex $\tilde{u} \in \mathrm{fib}_u$ we label the vertex $c(\tilde{u}) \in \mathrm{fib}_u$, $c \in \Gamma$, by (u, c). This way we obtain a bijective labeling of fib_u by $\{u\} \times \Gamma$. Similarly, the darts in fib_x are labeled by $\{x\} \times \Gamma$, where (x, c) is the label of the dart in fib_x having its initial vertex labeled by (u, c). For $x \in D(X)$, let $\zeta_x \in \Gamma$ be such an element of the voltage group that $(\mathrm{end}(x), \zeta_x)$ is the label of the terminal vertex of the dart in fib_x labeled by $(x, 1)$. Then the terminal vertex of any dart in fib_x, say, labeled by (x, c), is labeled by $(\mathrm{end}(x), c\zeta_x)$. Clearly, $\zeta_{x^{-1}} = \zeta_x^{-1}$. The respective regular covering projection $\wp_\zeta \colon \mathrm{Cov}(\zeta) \to X$ is equivalent to $\wp$, a concept that we are now going to define.

Two covering projections $\wp \colon \tilde{X} \to X$ and $\wp' \colon \tilde{X}' \to X$ are *isomorphic* if there exists an automorphism $g \in \mathrm{Aut}\, X$ and an isomorphism $\tilde{g} \colon \tilde{X} \to \tilde{X}'$ such that the following diagram

$$\begin{array}{ccc} \tilde{X} & \xrightarrow{\tilde{g}} & \tilde{X}' \\ {\scriptstyle \wp}\downarrow & & \downarrow{\scriptstyle \wp'} \\ X & \xrightarrow{g} & X \end{array}$$

is commutative. If in the above diagram one can choose $g = \mathrm{id}$, then the projections are *equivalent*. Covering projections are usually studied up to equivalence, or possibly up to isomorphism (which is considerably more difficult).

A voltage assignment $\zeta \colon D(X) \to \Gamma$ can be naturally extended to walks as follows: if $W = u_0\, x_1\, u_1\, x_2\, u_2\, \ldots\, u_{n-1}\, x_n\, u_n$, then $\zeta_W = \zeta_{x_1}\zeta_{x_2} \ldots \zeta_{x_n}$. Clearly, homotopic walks carry the same voltage, and so voltages can be assigned to homotopy classes. Moreover, the 'right action' of homotopy classes via unique path-lifting along $\wp_\zeta \colon \mathrm{Cov}(\zeta) \to X$ is essentially the voltage-action: if $W \colon u \to v$ is a walk in X and $\tilde{u} \in \mathrm{fib}_u$ is labeled by (u, c), then $\tilde{u} \cdot [W] \in \mathrm{fib}_v$ is labeled by $(v, c\,\zeta_W)$. We may therefore say that the voltage-action faithfully represents the 'action' of homotopy classes, and in particular, the action of $\pi(X, u)$. It immediately follows that ζ defines a group homomorphism $\zeta \colon \pi(X, u) \to \Gamma$ (denoted by the same symbol for convenience).

Lifts of automorphisms. An automorphism $g \in \mathrm{Aut}\, X$ *lifts along* a covering projection $\wp \colon \tilde{X} \to X$ if there exists an automorphism $\tilde{g} \in \mathrm{Aut}\, \tilde{X}$ such that the diagram

$$\begin{array}{ccc} \tilde{X} & \xrightarrow{\tilde{g}} & \tilde{X} \\ {\scriptstyle \wp}\downarrow & & \downarrow{\scriptstyle \wp} \\ X & \xrightarrow{g} & X \end{array}$$

is commutative. The automorphism $\tilde{g}$ then *projects* to g. A group $G \leq \mathrm{Aut}\, X$ lifts if all $g \in G$ lift. We call such a covering projection *G-admissible*. The collection of all lifts of all elements in G form a subgroup $\tilde{G} \leq \mathrm{Aut}\, \tilde{X}$, the *lift* of G. In particular, the lift of the trivial group is known as the *group of covering transformations* (or self-equivalences of $\wp$) and denoted by $\mathrm{CT}_\wp$. Moreover, the sequence

$$\mathrm{id} \to \mathrm{CT}_\wp \to \tilde{G} \to G \to \mathrm{id}$$

is short exact. In other words, $\tilde{G}$ is an extension of $\mathrm{CT}_\wp$ by G, and hence $g \in G$ has exactly $|\mathrm{CT}_\wp|$ distinct lifts, a coset of $\mathrm{CT}_\wp$ within $\tilde{G}$. Furthermore, if G lifts along a given projection $\wp$, then it lifts along any covering projection equivalent to $\wp$. This allows us to study lifts of automorphisms combinatorially in terms of voltages, see for instance [1, 21, 32, 33, 48, 50]. Also note that if $\tilde{G}$ and $\tilde{G}'$ are the lifts of G along equivalent projections $\wp$ and $\wp'$, respectively, then the short exact sequences $\mathrm{id} \to \mathrm{CT}_\wp \to \tilde{G} \to G \to \mathrm{id}$ and $\mathrm{id} \to \mathrm{CT}_{\wp'} \to \tilde{G}' \to G \to \mathrm{id}$ are isomorphic. Thus, structural properties of lifted groups can be studied combinatorially in terms of voltages as well. In this paper we focus on G-admissible covering projections such that the extension $\mathrm{id} \to \mathrm{CT}_\wp \to \tilde{G} \to G \to \mathrm{id}$ is split. We call such a covering projection *G-split-admissible*. Note, however, that the lifted group $\tilde{G}$ might contain a subgroup H isomorphic to $\mathrm{CT}_\wp$ such that the extension $\mathrm{id} \to H \to \tilde{G} \to G \to \mathrm{id}$ is split even if $\mathrm{id} \to \mathrm{CT}_\wp \to \tilde{G} \to G \to \mathrm{id}$ is not.

From now on we shall be assuming that covering projections are regular and moreover, that covering graphs are connected as well. By assuming connectedness we essentially do not loose on generality, however, we technically gain a lot. Let us remark that if X is connected, then the covering graph is connected if and only if the fundamental group $\pi(X, u)$ acts transitively on fib_u. An equivalent requirement in terms of a voltage assignment $\zeta\colon D(X) \to \Gamma$ is that the homomorphic image $\zeta(\pi(X,u)) \leq \Gamma$ acts transitively relative to its voltage-action on Γ, that is, $\zeta(\pi(X,u)) = \Gamma$, which in turn amounts to saying that the voltages assigned to fundamental closed walks at u generate the voltage group Γ. With the assumption on connectedness, the group $\mathrm{CT}_\wp$ acts regularly on each fibre, and hence any lift $\tilde{g}$ of $g \in \mathrm{Aut}\, X$, if it exists, is uniquely determined by the mapping of just one vertex (or dart). Also, the semiregular group C from the definition of a regular covering is now $C = \mathrm{CT}_\wp$, and the voltage assignment $\zeta\colon D(X) \to \Gamma$ that reconstructs the projection takes values in an abstract group $\Gamma \cong \mathrm{CT}_\wp$.

Consider now a regular covering projection $\wp_\zeta\colon \mathrm{Cov}(\zeta) \to X$ of connected graphs, where X is assumed to be finite, and let $u_0 \in V(X)$ be an arbitrarily chosen base vertex. By the *basic lifting lemma*, see [32, Theorem 4.2] and [33, Theorem 7.1], a group $G \leq \mathrm{Aut}\, X$ lifts along $\wp_\zeta$ if and only if any closed walk W at u_0 with $\zeta_W = 1$ is mapped to a walk with $\zeta_{gW} = 1$, for all $g \in G$. This is equivalent to requiring that for each $g \in G$ there exists an *induced automorphism* $g^{\#_{u_0}} \in \mathrm{Aut}\,\Gamma$ of the voltage group defined locally at u_0 by

$$g^{\#_{u_0}}(\zeta_W) = \zeta_{gW}, \qquad W \in \pi(X, u_0).$$

Note that if the condition is satisfied at u_0, it holds locally at any vertex. In general, for $g, h \in G$ the automorphisms $g^{\#_u}$ and $g^{\#_v}$ at distinct vertices, as well as the automorphisms $(gh)^{\#_u}$ and $g^{\#_u}h^{\#_u}$, differ by an inner automorphism of Γ. More precisely, the following holds. (Throughout the paper, Ψ_t denotes the inner automorphism $\Psi_t(a) = tat^{-1}$, whatever the group. Note further that all automorphisms are composed as functions.)

Proposition 2.1. *Let $G \leq \mathrm{Aut}\, X$ be a group of automorphisms that lifts along a regular covering projection of connected graphs $\wp\colon \tilde{X} \to X$ given in terms of a voltage assignment $\zeta\colon D(X) \to \Gamma$. Then for any $g, h \in G$ we have*

$$\begin{aligned} \Psi_{g^{\#_u}(\zeta_Q)\zeta_{gQ}^{-1}}\, g^{\#_v} &= g^{\#_u}, \quad Q\colon v \to u, \\ \Psi_{g^{\#_u}(\zeta_Q)\zeta_{gQ}^{-1}}\, (gh)^{\#_u} &= g^{\#_u}h^{\#_u}, \quad Q\colon hu \to u. \end{aligned}$$

Proof. Let W be a closed walk at v and $Q\colon v \to u$ an arbitrary walk. Then $Q^{-1}WQ$ is a closed walk at u, and by the definition of induced automorphisms of Γ at v and u we have $g^{\#_v}(\zeta_W) = \zeta_{gW}$ and $g^{\#_u}(\zeta_{Q^{-1}WQ}) = \zeta_{g(Q^{-1}WQ)}$. Clearly, $\zeta_{Q^{-1}WQ} = \zeta_Q^{-1}\zeta_W\zeta_Q$ and $\zeta_{g(Q^{-1}WQ)} = \zeta_{gQ}^{-1}\zeta_{gW}\zeta_{gQ}$. Since $g^{\#_u}$ is an automorphism we have

$$g^{\#_u}(\zeta_Q)^{-1}g^{\#_u}(\zeta_W)g^{\#_u}(\zeta_Q) = \zeta_{gQ}^{-1}\zeta_{gW}\zeta_{gQ}.$$

Hence $\Psi_{g^{\#_u}(\zeta_Q)\zeta_{gQ}^{-1}}\, g^{\#_v} = g^{\#_u}$, and the first part is proved. For the second part, let W be a closed walk at u and $Q\colon hu \to u$ an arbitrary walk. Then

$$(gh)^{\#_u}(\zeta_W) = \zeta_{ghW} = g^{\#_{hu}}(\zeta_{hW}) = g^{\#_{hu}}(h^{\#_u}(\zeta_W)).$$

Hence $(gh)^{\#_u} = g^{\#_{hu}}h^{\#_u}$. By the first part we have $\Psi_{g^{\#_u}(\zeta_Q)\zeta_{gQ}^{-1}}\, g^{\#_{hu}} = g^{\#_u}$, and consequently, $\Psi_{g^{\#_u}(\zeta_Q)\zeta_{gQ}^{-1}}\,(gh)^{\#_u} = g^{\#_u}h^{\#_u}$, as required. □

Clearly, the function

$$\#_{u_0}\colon G \to \operatorname{Aut}\Gamma, \quad g \mapsto g^{\#_{u_0}},$$

is not a group homomorphism in general. But if we define $g^{\#} = g^{\#_{u_0}} \bmod \operatorname{Inn}\Gamma$, then, by Proposition 2.1, $g^{\#}$ does not depend on u_0, and $\#\colon G \to \operatorname{Out}\Gamma$, $g \mapsto g^{\#}$, is a homomorphism. In particular, if the covering projection is *abelian*, meaning that $\Gamma \cong \mathrm{CT}_\wp$ is abelian, then $\# = \#_{u_0}\colon G \to \operatorname{Aut}\Gamma$ is a homomorphism, which turns Γ into a $\mathbb{Z}[G]$-module. We shall make substantial use of this fact later on.

If g lifts, denote by $\Phi_{v,\tilde g}$ the permutation on the voltage group Γ corresponding to the restriction $\tilde g\colon \mathrm{fib}_v \to \mathrm{fib}_{gv}$. In other words,

$$\tilde g(v,c) = (gv, \Phi_{v,\tilde g}(c)). \tag{2.1}$$

As it was shown in [32, 33], the mappings of labels at different fibres relate to each other as follows:

$$\Phi_{u,\tilde g}(c) = \Phi_{u,\tilde g}(1)\, g^{\#_u}(c) \tag{2.2}$$
$$\Phi_{v,\tilde g}(c) = \Phi_{u,\tilde g}(c)\, g^{\#_u}(\zeta_Q)\zeta_{gQ}^{-1}, \tag{2.3}$$

where $Q\colon v \to u$ is an arbitrary walk. Finally, for $t \in \Gamma$ we denote by $\tilde g_t$ the uniquely defined lift of g mapping the vertex in fib_{u_0} labeled by $1 \in \Gamma$ to the vertex in fib_{gu_0} labeled by $t \in \Gamma$, that is,

$$\tilde g_t(u_0, 1) = (gu_0, t).$$

In particular, $\tilde{\mathrm{id}}_t$ is the covering transformation acting on the second coordinates in $\operatorname{Cov}(\zeta)$ by left multiplication by t on Γ. Indeed, since $\mathrm{id}^{\#_u} = \mathrm{id}$ for all $u \in V(X)$, it follows from (2.2) and (2.3) that

$$\tilde{\mathrm{id}}_t(u,c) = (u, tc).$$

3 Extensions in terms of voltages

The method how to recapture a given group extension $1 \to K \to E \to Q \to 1$ in the form of a crossed product is known and goes back to Schreier (cf. [38]). First choose a *system of coset representatives* of K within E (also called an *algebraic transversal*) $T = \{t_x \mid x \in Q\}$ (and usually *normalized* in the sense that $t_1 = 1$). Then compute the *factor set* $\mathcal{F}\colon Q \times Q \to K$ defined by

$$\mathcal{F}\colon (x, y) \mapsto t_x t_y t_{xy}^{-1},$$

and the function $\Psi\colon Q \to \operatorname{Aut} K$, $x \mapsto \Psi_{t_x}$ (recall that $\Psi_{t_x}(a) = t_x a t_x^{-1}$); in general, Ψ is not a group homomorphism, and is often referred to as the *weak action* of Q on K (which, when reduced modulo inner automorphisms of K, gives rise to a homomorphism $Q \to \operatorname{Out} K$ known as the *coupling* or the *twisting map*). These data determine a group operation on $K \times Q$ defined by

$$(a, x)(b, y) = (a\Psi_{t_x}(b)\mathcal{F}(x, y), xy).$$

The resulting group is called the *crossed product* of K by Q and denoted $K \times_{\Psi,\mathcal{F}} Q$. The mapping $K \times_{\Psi,\mathcal{F}} Q \to E$ defined by $(a, x) \mapsto at_x$ is an isomorphism taking $K \times 1$ onto K and $1 \times Q$ onto the algebraic transversal T, and establishes an *equivalence* of short exact sequences

$$\begin{array}{ccccccccc} & & & \nearrow & K \times_{\Psi,\mathcal{F}} Q & \searrow & & & \\ 1 & \longrightarrow & K & & \downarrow & & Q & \longrightarrow & 1. \\ & & & \searrow & E & \nearrow & & & \end{array}$$

Suppose now that a regular covering projection $\wp = \wp_\zeta\colon \operatorname{Cov}(\zeta) \to X$ of connected graphs is given in terms of a Cayley voltage assignment $\zeta\colon D(X) \to \Gamma$, and let a group $G \le \operatorname{Aut} X$ of automorphisms lift to $\tilde{G} \le \operatorname{Aut}\operatorname{Cov}(\zeta)$.

In order to recapture $\tilde{G}$ as a crossed product of $\mathrm{CT}_\wp$ by G let us choose a particular algebraic transversal by taking $t_g = 1$ for all $g \in G$, that is, $T = \{\tilde{g}_1 \mid g \in G\}$. To compute the factor set $G \times G \to \mathrm{CT}_\wp$ we need to identify $c \in \Gamma$ such that $\tilde{g}_1 \tilde{h}_1 (\widetilde{gh})_1^{-1} = \tilde{\mathrm{id}}_c$. We do that by evaluating $\tilde{g}_1 \tilde{h}_1 = \tilde{\mathrm{id}}_c (\widetilde{gh})_1$ at $(u_0, 1)$. Using (2.2) and (2.3) we get

$$c = g^{\#_{u_0}}(\zeta_Q)\zeta_{gQ}^{-1}, \quad \text{where} \quad Q\colon hu_0 \to u_0$$

is arbitrary. As for the weak action $G \to \operatorname{Aut}\mathrm{CT}_\wp$ we need to find $t \in \Gamma$ such that $\tilde{g}_1 \tilde{\mathrm{id}}_a \tilde{g}_1^{-1} = \tilde{\mathrm{id}}_t$. Evaluating $\tilde{g}_1 \tilde{\mathrm{id}}_a = \tilde{\mathrm{id}}_t \tilde{g}_1$ at $(u_0, 1)$ and using (2.2) we obtain

$$t = g^{\#_{u_0}}(a).$$

In view of the isomorphism $\mathrm{CT}_\wp \cong \Gamma$, $\tilde{\mathrm{id}}_t \mapsto t$, we have an isomorphism of short exact sequences

$$\begin{array}{ccccccccc} \mathrm{id} & \longrightarrow & \mathrm{CT}_\wp & \longrightarrow & \tilde{G} & \longrightarrow & G & \longrightarrow & \mathrm{id} \\ \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\ 1 & \longrightarrow & \Gamma & \longrightarrow & \tilde{G} & \longrightarrow & G & \longrightarrow & \mathrm{id}. \end{array}$$

Hence the lifted group $\tilde{G}$ can be written, up to isomorphism, as a crossed product $\Gamma \times_{\Psi,\mathcal{F}} G$, where $\mathcal{F}\colon G \times G \to \Gamma$ is given by $\mathcal{F}(g,h) = g^{\#_{u_0}}(\zeta_Q)\zeta_{gQ}^{-1}$ and $\Psi\colon G \to \operatorname{Aut}\Gamma$ is defined by $\Psi_g = g^{\#_{u_0}}$. Note that the weak action Ψ is precisely $\#_{u_0}$ defined in Preliminaries. We have therefore proved the following theorem.

Theorem 3.1. *Let $\wp = \wp_\zeta\colon \operatorname{Cov}(\zeta) \to X$ be a regular covering projection of connected graphs given in terms of a Cayley voltage assignment $\zeta\colon D(X) \to \Gamma$, and let a group $G \leq \operatorname{Aut} X$ of automorphisms lift to $\tilde{G} \leq \operatorname{Aut}\operatorname{Cov}(\zeta)$. Choosing a base vertex $u_0 \in V(X)$, let $\Psi\colon G \to \operatorname{Aut}\Gamma$ and $\mathcal{F}\colon G \times G \to \Gamma$ be functions defined by*

$$\Psi_g = g^{\#_{u_0}} \quad \textit{and} \quad \mathcal{F}(g,h) = g^{\#_{u_0}}(\zeta_Q)\zeta_{gQ}^{-1},\ \ Q\colon hu_0 \to u_0,$$

respectively. Then there is an isomorphism

$$\Gamma \times_{\Psi,\mathcal{F}} G \to \tilde{G}, \qquad (a,g) \mapsto \tilde{g}_a$$

taking $\Gamma \times \mathrm{id}$ onto $\mathrm{CT}_\wp$ and $1 \times G$ onto the algebraic transversal $\{\tilde{g}_1 \mid g \in G\}$. □

4 Split extensions in terms of voltages

Recall that a short exact sequence $1 \to K \to E \to Q \to 1$ is *split* if there exists an algebraic transversal $T = \{t_x \mid x \in Q\}$ which is a subgroup, called a *complement* to K within E. Relative to such a complement, the respective factor set $\mathcal{F} \equiv 1$ is trivial and the weak action Ψ is in fact an action, that is, $\Psi\colon Q \to \operatorname{Aut} K$ is a homomorphism. Consequently, recapturing E as the corresponding crossed product results in a *semidirect product* $K \rtimes_\Psi Q$ with the group operation $(a,x)(b,y) = (a\Psi_{t_x}(b), xy)$.

In the next theorem, the necessary and sufficient condition for a regular covering projection $\wp$ to be G-split-admissible, together with an explicit description of the lifted group as a semidirect product of $\mathrm{CT}_\wp$ by G, are given in terms of voltages.

Theorem 4.1. *Let $\wp = \wp_\zeta\colon \operatorname{Cov}(\zeta) \to X$ be a regular covering projection of connected graphs given in terms of a Cayley voltage assignment $\zeta\colon D(X) \to \Gamma$, and let a group $G \leq \operatorname{Aut} X$ of automorphisms lift to $\tilde{G} \leq \operatorname{Aut}\operatorname{Cov}(\zeta)$. Then $\wp$ is G-split admissible if and only if there exists a normalized function $t\colon G \to \Gamma$ (that is, $t_{\mathrm{id}} = 1$) such that*

$$t_{gh} = t_g g^{\#_{u_0}}(t_h)\, g^{\#_{u_0}}(\zeta_Q)\zeta_{gQ}^{-1} \tag{4.1}$$

where $Q\colon hu_0 \to u_0$ is an arbitrary walk. In this case there exists a homomorphism $\theta\colon G \to \operatorname{Aut}\Gamma$ given by

$$\theta_g(c) = t_g g^{\#_{u_0}}(c) t_g^{-1}, \tag{4.2}$$

and $(a,g) \mapsto \tilde{g}_{a\,t_g}$ defines an isomorphism $\Gamma \rtimes_\theta G \to \tilde{G}$ which takes $\Gamma \times \mathrm{id}$ onto $\mathrm{CT}_\wp$ and $\mathrm{id} \times G$ onto the algebraic transversal $\overline{G} = \{\tilde{g}_{t_g} \mid g \in G\}$, a complement to $\mathrm{CT}_\wp$.

Proof. Let us recover the lifted group $\tilde{G}$ as in Theorem 3.1. The extension splits if and only if there exists an algebraic transversal $\{(t_g, g),\ g \in G\}$ to $\Gamma \times \mathrm{id}$ in $\Gamma \times_{\Psi,\mathcal{F}} G$ which is a subgroup. Equivalently, we must have $(t_{gh}, gh) = (t_g, g)(t_h, h)$. By the definition of multiplication in $\Gamma \times_{\Psi,\mathcal{F}} G$ the right hand side is equal to $(t_g g^{\#_{u_0}}(t_h)\mathcal{F}(g,h), gh)$. Hence the necessary and sufficient condition (4.1) can be expressed as stated in the theorem.

That (4.2) defines a homomorphism can be shown by computation, using (4.1) and Proposition 2.1. The rest is straightforward as well. □

Remark 4.2. In the abelian case, (4.1) rewrites as $t_{gh} = t_g + g^{\#_{u_0}}(t_h) + \tau(g,h)$, where $\tau(g,h) = \mathcal{F}(g,h) = g^{\#_{u_0}}(\zeta_Q) - \zeta_{gQ}$ is the 2-cocycle. Thus, (4.1) is equivalent to the fact that $\tau(g,h) = t_{gh} - t_g - g^{\#_{u_0}}(t_h)$ must be a 2-coboundary. □

From Theorem 4.1 we readily obtain the necessary and sufficient conditions for G to lift as a *direct product extension* of $\mathrm{CT}_\wp$, that is, when $\mathrm{CT}_\wp$ has a normal complement within the lifted group $\tilde{G}$.

Theorem 4.3. *Let $\wp = \wp_\zeta \colon \mathrm{Cov}(\zeta) \to X$ be a regular covering projection of connected graphs given in terms of a Cayley voltage assignment $\zeta \colon D(X) \to \Gamma$. Then G lifts along $\wp$ as a direct product extension of $\mathrm{CT}_\wp$ if and only if there exists a normalized function $t \colon G \to \Gamma$ (that is, $t_{\mathrm{id}} = 1$) satisfying*

$$t_{gh} = t_h \zeta_Q t_g \zeta_{gQ}^{-1}, \tag{4.3}$$

where $Q \colon hu_0 \to u_0$ is an arbitrary walk. In this case, $(a,g) \mapsto \tilde{g}_{a\,t_g}$ defines an isomorphism $\Gamma \times G \to \tilde{G}$ which takes $\Gamma \times \mathrm{id}$ onto $\mathrm{CT}_\wp$ and $\mathrm{id} \times G$ onto the algebraic transversal $\overline{G} = \{\tilde{g}_{t_g} \mid g \in G\}$, a normal complement to $\mathrm{CT}_\wp$.

Proof. Suppose that G lifts as a direct product such that $\mathrm{CT}_\wp$ has a normal complement $\overline{G} = \{\tilde{g}_{t_g} \mid g \in G\}$. By Theorem 4.1 the respective function $t \colon G \to \Gamma$ satisfies (4.1). Normality of $\overline{G}$ implies that $\theta_g(c)$ given by (4.2) must be the identity automorphism. Hence $g^{\#_{u_0}}(c) = t_g^{-1} c t_g$, and by (4.1) we have $t_{gh} = t_h \zeta_Q t_g \zeta_{gQ}^{-1}$, as required.

For the converse suppose that a function $t \colon G \to \Gamma$ satisfies $t_{gh} = t_h \zeta_Q t_g \zeta_{gQ}^{-1}$. Taking $h = \mathrm{id}$ we obtain $\zeta_{gQ} = t_g^{-1} \zeta_Q t_g$ for all closed walks $Q \colon u_0 \to u_0$. Therefore, if $\zeta_Q = 1$ then $\zeta_{gQ} = 1$ for all $g \in G$. By the basic lifting lemma G lifts, and $g^{\#_{u_0}}$ takes the form $g^{\#_{u_0}}(c) = t_g^{-1} c t_g$. It follows that $t_{gh} = t_g g^{\#_{u_0}}(t_h)\, g^{\#_{u_0}}(\zeta_Q) \zeta_{gQ}^{-1}$. By Theorem 4.1 we have $\tilde{G} \cong \Gamma \rtimes_\theta G$ where $\theta_g(c) = t_g g^{\#_{u_0}}(c) t_g^{-1} = c$. Hence $\tilde{G} \cong \Gamma \times G$, and the proof is complete. □

Remark 4.4. Notice the subtle difference in assumptions in Theorems 4.1 and 4.3. While in 4.1 we had to assume in advance that G had a lift, this assumption is not required in 4.3 as condition (4.3) does not involve $g^{\#_{u_0}}$. □

We also note the following. Suppose that G lifts as a split extension of $\mathrm{CT}_\wp$. In general, normal and non-normal complements to $\mathrm{CT}_\wp$ might exist. So a priori knowledge about a given extension being split does not make it easier to check whether the extension is actually a direct product extension. In the abelian case, however, things are different since complements are either all normal or all non-normal. This means that if $t \colon G \to \mathrm{Aut}\,\Gamma$ is just any normalized function satisfying (4.1), the extension will be a direct product extension if and only if the corresponding homomorphism θ as in (4.2) is trivial.

For later reference, see Corollary 5.5, we explicitly record the following corollary.

Corollary 4.5. *Let $\wp = \wp_\zeta \colon \mathrm{Cov}(\zeta) \to X$ be a regular covering projection of connected graphs given in terms of a Cayley voltage assignment $\zeta \colon D(X) \to \Gamma$. Suppose that $G \leq \mathrm{Aut}\,X$ lifts along $\wp$ as a split extension. Then G lifts as a direct product extension of $\mathrm{CT}_\wp$ if and only if $\zeta_{gW} = \zeta_W$ holds for all closed walks W from a basis of the first homology group $H_1(X,\mathbb{Z})$ and all g from some generating set of G.*

Proof. By Theorem 4.1 there exists a normalized function $t\colon G \to \Gamma$ satisfying (4.1). Now, in the abelian case the extension is a direct product extension if and only if θ as in (4.2) is trivial. Since $\theta_g(c) = g^{\#_{u_0}}(c)$, this amounts to saying that $\zeta_{gW} = \zeta_W$ must hold for all closed walks based at u_0 and all $g \in G$. Moreover, recall that in the abelian case $g^{\#_{u_0}}$ does not depend on u_0. Hence the above necessary and sufficient condition can be replaced by only considering closed walks from a basis of the first homology group. Clearly, it is enough to consider just the automorphisms from a generating set of G. □

Remark 4.6. Note that if the covering projection is abelian and the condition $\zeta_{gW} = \zeta_W$ holds true for all closed walks and all $g \in G$, then G clearly lifts (by the basic lifting lemma). However, the extension might not be split.

As an example, let X be the 2-dipole with vertices 1 and 2 and two parallel links from 1 to 2 defined by the darts a and b. The voltage assignment $\zeta_a = \zeta_{a^{-1}} = 0$, $\zeta_b = \zeta_{b^{-1}} = 1$ in the group $\mathbb{Z}_2$ gives rise to a connected covering graph isomorphic to the 4-cycle C_4. Clearly, $\zeta_{gW} = \zeta_W$ holds for all $g \in \operatorname{Aut} X \cong \mathbb{Z}_2 \times \mathbb{Z}_2$ and all closed walks W. However, the lifted group is isomorphic to D_4, viewed as a central extension of $\mathbb{Z}_2$ by $\mathbb{Z}_2 \times \mathbb{Z}_2$, and this extension is clearly not split. □

5 Algorithmic aspects

Let $\wp = \wp_\zeta\colon \operatorname{Cov}(\zeta) \to X$ be a regular covering projection of connected graphs given in terms of a Cayley voltage assignment $\zeta\colon D(X) \to \Gamma$, where X is assumed to be finite, and let $G \le \operatorname{Aut} X$ be a group of automorphisms. Speaking of algorithmic and complexity issues related to lifting automorphisms one would certainly first need to address the question of how difficult is to test whether G lifts at all. However, this will not be our concern here; the problem has been considered, to some extent, in [48].

Assuming that G is known to have a lift we focus on efficient algorithms (in terms of voltages) for testing if G lifts as a split extension of $\mathrm{CT}_\wp$. Testing condition (4.1) of Theorem 4.1 is hard even if Γ is abelian – as one has to take into account all group elements of G. (Indeed, Theorem 4.1 is of purely theoretical interest.) A much better alternative would be to consider just the generators, and in fact, one must then assume that G is given by a presentation, which is sensible assumption. Proposition 5.1 below is a reformulation of a standard result, c.f. [23, Lemma 2.76], tailored to our present needs. For completeness we provide the proof.

Proposition 5.1. *Let $\wp\colon \tilde{X} \to X$ be a regular covering projection of connected graphs, and let $G \le \operatorname{Aut} X$ be a group given by the presentation $G = \langle S \mid R\rangle$, where $S = \{g_1, g_2, \ldots, g_n\}$ and the R-relations are $R_j(g_1, g_2, \ldots, g_n) = \mathrm{id}$, $j = 1, 2, \ldots, m$. Suppose that G lifts. Then the lifted group $\tilde{G}$ is a split extension of $\mathrm{CT}_\wp$ if and only if there are lifts $\bar{g}_1, \bar{g}_2, \ldots, \bar{g}_n$ of $g_1, g_2, \ldots, g_n$, respectively, satisfying the defining relations $R_j(\bar{g}_1, \bar{g}_2, \ldots, \bar{g}_n) = \mathrm{id}$, $j = 1, 2, \ldots, m$.*

Proof. Suppose first that there are lifts $\bar{g}_1, \bar{g}_2, \ldots, \bar{g}_n$ of $g_1, g_2, \ldots, g_n$ satisfying the R-relations, and let $C = \langle \bar{g}_1, \bar{g}_2, \ldots, \bar{g}_n\rangle \le \tilde{G}$. Since the R-relations are the defining relations of G there exists an epimorphism $G \to C$, $g_i \mapsto \bar{g}_i$. On the other hand, C projects onto G, with $\bar{g}_i \mapsto g_i$. Consequently, $C \cong G$. As C isomorphically projects onto G it must intersect the kernel $\mathrm{CT}_\wp$ of the projection $\tilde{G} \to G$ trivially. Hence C is a complement to $\mathrm{CT}_\wp$ within $\tilde{G}$.

Conversely, suppose that there are lifts $\bar{g}_1, \bar{g}_2, \ldots, \bar{g}_n$ of $g_1, g_2, \ldots, g_n$ such that $C = \langle \bar{g}_1, \bar{g}_2, \ldots, \bar{g}_n \rangle \leq \tilde{G}$ is a complement to $\mathrm{CT}_\wp$ within $\tilde{G}$. Then $C \cong G$, and since $\bar{g}_i \mapsto g_i$ we have that each automorphism $R_j(\bar{g}_1, \bar{g}_2, \ldots, \bar{g}_n)$ projects to $R_j(g_1, g_2, \ldots, g_n) = \mathrm{id}$. So $R_j(\bar{g}_1, \bar{g}_2, \ldots, \bar{g}_n) \in C$ belongs to $\mathrm{CT}_\wp$. As C is the complement we have $R_j(\bar{g}_1, \bar{g}_2, \ldots, \bar{g}_n) = \mathrm{id}$, and the proof is complete. □

The condition $R_j(\bar{g}_1, \bar{g}_2, \ldots, \bar{g}_n) = \mathrm{id}$ can be tested just by checking whether the automorphism $R_j(\bar{g}_1, \bar{g}_2, \ldots, \bar{g}_n)$, which necessarily belongs to $\mathrm{CT}_\wp$, fixes a vertex. With our assumption that the covering graph is reconstructed as $\mathrm{Cov}(\zeta)$ we choose this vertex to be $(u_0, 1)$. Let $\bar{g}_i(u_0, 1) = (g_i u_0, t_i)$, and recall that a lift is uniquely determined by the image of a single vertex. If $t_1, t_2, \ldots, t_n$ are explicitly given, then $R_j(\bar{g}_1, \bar{g}_2, \ldots, \bar{g}_n)$ can be evaluated recursively using (2.2) and (2.3). To find whether the required lifts $\bar{g}_1, \bar{g}_2, \ldots, \bar{g}_n$ exist by checking the whole set Γ^n for all possible values of $t_1, t_2, \ldots, t_n$ is far from optimal. The core of the problem is therefore to evaluate $R_j(\bar{g}_1, \bar{g}_2, \ldots, \bar{g}_n)$ efficiently when $t_1, t_2, \ldots, t_n$ are seen as symbolic variables, in which case the requirements $R_j(\bar{g}_1, \bar{g}_2, \ldots, \bar{g}_n)(u_0, 1) = (u_0, 1)$ translate to an equivalent problem of solving a system of equations in the variables $t_1, t_2, \ldots, t_n \in \Gamma$.

We are faced with two main difficulties. First, to evaluate $R_j(\bar{g}_1, \bar{g}_2, \ldots, \bar{g}_n)$ using symbolic variables we need to express $g^{\#_{u_0}}$ by a 'closed formula', and second, we have to solve a (possibly a non-linear) system of equations over Γ. Both are rather hopeless if Γ is nonabelian. On the other hand, if $\Gamma \cong \mathrm{CT}_\wp$ is a finitely presented abelian group, then the automorphisms of Γ can be represented by integer matrices acting on the left on integer column vectors representing group elements. (In what follows, we shall be freely using the term 'vector' for ease of expression.) Moreover, as we shall see, in the abelian case the system of equations results in a linear system over the integers.

5.1 Abelian covers

Let us therefore assume that Γ is abelian, given by a presentation $\Gamma = \langle \Delta \mid \Lambda \rangle$, where $\Delta = \{c_1, c_2, \ldots, c_r\}$ is a generating set and $\Lambda_k(c_1, c_2, \ldots, c_r) = 0$, where $k = 1, 2, \ldots, s$, are the Λ-relations. Each element $c \in \Gamma$ can be represented by a column vector $\underline{c} \in \mathbb{Z}^{r,1}$ such that

$$\underline{c} = [\lambda_1, \lambda_2, \ldots, \lambda_r]^T, \quad \text{where} \quad c = \sum_{i=1}^{r} \lambda_i c_i.$$

This representation is unique modulo the kernel (generated by the defining relations Λ_j) of the natural quotient projection $\kappa\colon \mathbb{Z}^{r,1} \to \Gamma$. Moreover, any automorphism $\phi \in \mathrm{Aut}\,\Gamma$ can be represented (again not in a unique way) as a matrix over $\mathbb{Z}$ by expressing each $\phi(c_i)$ as $\phi(c_i) = \sum_{j=1}^{r} \alpha_{ji}\, c_j$, and taking $M_\phi = [\alpha_{ij}] \in \mathbb{Z}^{r,r}$. Clearly, the following diagram

$$\begin{array}{ccc} \mathbb{Z}^{r,1} & \xrightarrow{M_\phi} & \mathbb{Z}^{r,1} \\ \kappa\downarrow & & \downarrow\kappa \\ \Gamma & \xrightarrow[\phi]{} & \Gamma \end{array}$$

is commutative, or in other words, evaluation of the automorphism ϕ is given by $\phi(c) = \kappa(M_\phi \underline{c})$.

Coming back to our original setting of evaluating the lifted automorphisms, recall from Preliminaries that in the abelian case the automorphism $g^{\#} = g^{\#_{u_0}}$ does not depend on the base vertex, and that $(gh)^{\#} = g^{\#}h^{\#}$. Also, we shall simplify the notation for the matrix representing $g_i^{\#}$ by writing $M_i = M_{g_i^{\#}}$. In view of (2.2) and (2.3), the formula for evaluating the lifted automorphism $\bar{g}_i$ at an arbitrary vertex (v, c) is now given by

$$\Phi_{v,\bar{g}_i}(c) = t_i + g_i^{\#}(c) + g_i^{\#}(\zeta_Q) - \zeta_{g_i Q},$$

where $Q\colon v \to u_0$ is an arbitrary walk. This can be rewritten in vector form as

$$\underline{\Phi_{v,\bar{g}_i}(c)} = \underline{t_i} + M_i \underline{c} + M_i \underline{\zeta_Q} - \underline{\zeta_{g_i Q}}. \tag{5.1}$$

So far we have overcome part of the problem: representation of $g_i^{\#}$'s by a 'closed formula'. However, since $\underline{t_i}$'s and the vector $\underline{c}$ (which linearly depends on t_i's when the formula is applied recursively while processing R_j) are symbolic variables, the evaluation requires symbolic computation – and that is something we still want to avoid. To this end we do the following.

Let $\mathbf{t} = [\underline{t_1}^T, \underline{t_2}^T, \ldots, \underline{t_n}^T]^T \in \mathbb{Z}^{rn,1}$ be the 'extended' column of all the vectors $\underline{t_1}, \underline{t_2}, \ldots, \underline{t_n}$, and let $\mathbf{E}_i = [\mathbf{0}, \ldots, \mathbf{0}, \mathbf{I}, \mathbf{0}, \ldots, \mathbf{0}] \in \mathbb{Z}^{r,rn}$ be the matrix consisting of $n-1$ zero submatrices $\mathbf{0} \in \mathbb{Z}^{r,r}$ and one identity submatrix $\mathbf{I} \in \mathbb{Z}^{r,r}$ at 'i-th position'. Clearly, $\underline{t_i} = \mathbf{E}_i \mathbf{t}$. At each iteration step of the evaluation we can express the vector $\underline{c}$ linearly in terms of $\mathbf{t}$ as $\underline{c} = \mathbf{A}_j \mathbf{t} + \underline{b}_j$, for an appropriate matrix $\mathbf{A}_j \in \mathbb{Z}^{r,rn}$ and vector $\underline{b}_j \in \mathbb{Z}^{r,1}$, neither of which depends on $\mathbf{t}$.

Indeed. Suppose that while scanning the relator R_j from right to left we need to evaluate $\bar{g}_i$ at vertex $(v, \kappa \underline{c})$. Substituting $\underline{c}$ with $\mathbf{A}_j \mathbf{t} + \underline{b}_j$ in (5.1) we get

$$\underline{\Phi_{v,\bar{g}_i}(c)} = (\mathbf{E}_i + M_i \mathbf{A}_j)\mathbf{t} + M_i(\underline{b}_j + \underline{\zeta_Q}) - \underline{\zeta_{g_i Q}}, \quad Q\colon v \to u_0,$$

and so the label $\Phi_{v,\bar{g}_i}(\kappa \underline{c})$ (the modified $\underline{c}$ as the input at the next step) is again of the form $\mathbf{A}_j \mathbf{t} + \underline{b}_j$, with $\mathbf{A}_j$ substituted by $\mathbf{E}_i + M_i \mathbf{A}_j$ and $\underline{b}_j$ substituted by $M_i(\underline{b}_j + \underline{\zeta_Q}) - \underline{\zeta_{g_i Q}}$. Initially, $\mathbf{A}_j$ is the zero matrix and $\underline{b}_j$ the zero vector. The method for evaluating the automorphism $R_j(\bar{g}_1, \bar{g}_2, \ldots, \bar{g}_n)$ is formally encoded in algorithm Evaluate.

Let the evaluation $R_j(\bar{g}_1, \bar{g}_2, \ldots, \bar{g}_n)(u_0, 0)$ using algorithm Evaluate terminate with $\underline{\Phi_{u_0,R_j}(0)} = \mathbf{A}_j \mathbf{t} + \underline{b}_j$, for $j = 1, 2, \ldots, m$. Then $R_j(\bar{g}_1, \bar{g}_2, \ldots, \bar{g}_n)(u_0, 0) = (u_0, 0)$ is equivalent to $\mathbf{A}_j \mathbf{t} + \underline{b}_j \in \operatorname{Ker} \kappa$, for all j. Putting together we must have

$$[\mathbf{A}_1^T, \mathbf{A}_2^T, \ldots, \mathbf{A}_m^T]^T \, \mathbf{t} = -[\underline{b}_1^T, \underline{b}_2^T, \ldots, \underline{b}_m^T]^T \tag{5.2}$$

modulo the relations Λ_j. Introducing additional auxiliary variables $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_m \in \mathbb{Z}^{s,1}$ we finally obtain the following linear system over $\mathbb{Z}$

$$\begin{bmatrix} \mathbf{A}_1 & \mathbf{\Lambda} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{A}_2 & \mathbf{0} & \mathbf{\Lambda} & \ldots & \mathbf{0} \\ \vdots & \vdots & & \ddots & \vdots \\ \mathbf{A}_m & \mathbf{0} & \ldots & & \mathbf{\Lambda} \end{bmatrix} \begin{bmatrix} \mathbf{t} \\ \underline{x}_1 \\ \underline{x}_2 \\ \vdots \\ \underline{x}_m \end{bmatrix} = - \begin{bmatrix} \underline{b}_1 \\ \underline{b}_2 \\ \vdots \\ \underline{b}_m \end{bmatrix}, \tag{5.3}$$

where $\mathbf{\Lambda} = [\underline{\Lambda_1}, \underline{\Lambda_2}, \ldots, \underline{\Lambda_s}] \in \mathbb{Z}^{r,s}$ and $\underline{\Lambda_k}$ is the vector representing the relator Λ_k.

Algorithm: Evaluate

Input: word $R_j(g_1, g_2, \ldots, g_n)$,
set T of $|V(X)|$ vectors $\underline{\zeta_Q} \in \mathbb{Z}^{r,1}$ with $Q : v \to u_0$ for all $v \in V(X)$,
list $\mathcal{Z}$ of n sets each containing $|V(X)|$ vectors $\underline{\zeta_{g_i Q}} \in \mathbb{Z}^{r,1}$,
list $\mathcal{M}$ of n matrices $M_i \in \mathbb{Z}^{r,r}$ representing $g_i^{\#}$

Output: matrix $\mathbf{A}_j \in \mathbb{Z}^{r,rn}$, vector $\underline{b}_j \in \mathbb{Z}^{r,1}$

set $\mathbf{A}_j \in \mathbb{Z}^{r,rn}$ to be the zero matrix and $\underline{b}_j \in \mathbb{Z}^{r,1}$ the zero vector; $v \leftarrow u_0$;
suppose $R_j = g_{k_l} \cdots g_{k_1}$;
for $i \leftarrow 1$ **to** l **do** (*scan word R_j from right to left*)
    $\mathbf{A}_j \leftarrow \mathcal{M}[k_i]\mathbf{A}_j$; (*multiply $\mathbf{A}_j$ on the left by M_{k_i}*)
    for $s \leftarrow 1$ **to** r **do** (*add $\mathbf{E}_{k_i}$ to $\mathbf{A}_j$*)
        $\mathbf{A}_j[s][r * k_i + s] \leftarrow \mathbf{A}_j[s][r * k_i + s] + 1$;
    let $\underline{\zeta_Q} \in T$ and $\underline{\zeta_{g_{k_i} Q}} \in \mathcal{Z}[k_i]$ with $Q : v \to u_0$;
    $\underline{b}_j \leftarrow \mathcal{M}[k_i](\underline{b}_j + \underline{\zeta_Q}) - \underline{\zeta_{g_{k_i} Q}}$;
    $v \leftarrow g_{k_i}(v)$;
return $\mathbf{A}_j$, $\underline{b}_j$

The problem of testing whether a given extension is split-admissible has now been reduced to an equivalent problem of checking whether the linear system (5.3) has a solution. Efficient algorithms for solving a system of linear equations over $\mathbb{Z}$ are long known, and are based on reducing the matrix of coefficients into Hermite or Smith normal form, see [23, Sections 9.2.3 and 9.2.4].

Theorem 5.2. *Let X be a finite connected graph and $G \leq \operatorname{Aut} X$ a group of automorphisms given by a presentation $G = \langle S \mid R \rangle$, where $S = \{g_1, g_2, \ldots, g_n\}$ is a generating set and $R_j(g_1, g_2, \ldots, g_n) = \mathrm{id}$, $j = 1, 2, \ldots, m$, are the R-relations. Further, let $\wp = \wp_\zeta \colon \operatorname{Cov}(\zeta) \to X$ be an abelian G-admissible regular covering projection of connected graphs arising from a Cayley voltage assignment $\zeta \colon D(X) \to \Gamma$. Suppose that the abelian group Γ is given by a presentation $\Gamma = \langle \Delta \mid \Lambda \rangle$, where $\Delta = \{c_1, c_2, \ldots, c_r\}$ is a generating set and $\Lambda_k(c_1, c_2, \ldots, c_r) = 0$, $k = 1, 2, \ldots, s$, are the Λ-relations.*

Then the short exact sequence $\mathrm{id} \to \mathrm{CT}_\wp \to \tilde{G} \to G \to \mathrm{id}$ is split if and only if the system of linear equations (5.3) has a solution in $\mathbb{Z}$. Moreover, let $\Omega \subseteq \Gamma^n$ be the set of all solutions of (5.3) reduced relative to the defining relations in Λ. Then Ω is in bijective correspondence with all complements to $\mathrm{CT}_\wp$ within $\tilde{G}$ (which correspond to all derivations $G \to \Gamma$), and two solutions in Ω correspond to conjugate complements if and only if they differ by an inner derivation.

Proof. In view of Proposition 5.1 and the above discussion it is clear that the extension splits if and only if the linear system (5.3) has an integer solution. It is also clear that each complement to $\mathrm{CT}_\wp$ in $\tilde{G}$ corresponds to some solution of (5.3), and hence to a solution in Ω. Moreover, two distinct solutions from Ω give rise to distinct complements. Indeed. Suppose that $(t_1, t_2, \ldots, t_n)$ and $(t'_1, t'_2, \ldots, t'_n)$ are two distinct solutions from Ω giving rise to the same complement $\overline{G}$. Then there is an index i such that $t_i \neq t'_i$, that is, $\bar{g}_i \neq \bar{g}'_i$. As $\bar{g}_i$ and $\bar{g}'_i$ are two lifts of the same automorphism g_i, we must have $\bar{g}'_i = \tilde{\mathrm{id}}_c \, \bar{g}_i$, where $\tilde{\mathrm{id}}_c \in \mathrm{CT}_\wp$. But since $\overline{G}$ is a complement to $\mathrm{CT}_\wp$ we must have $\bar{g}'_i = \bar{g}_i$, and therefore

$t_i = t_i'$, contrary to the assumption. It follows that all solutions in Ω correspond bijectively to all complements.

Let $\overline{G}$ and $\overline{G}'$ be two conjugate complements. Without loss of generality we may assume they are conjugate by an element $\tilde{\mathrm{id}}_c \in \mathrm{CT}_\wp$, that is, $\overline{G}' = \tilde{\mathrm{id}}_c \overline{G}\, \tilde{\mathrm{id}}_c^{-1}$. Since for any $\bar{g} \in \overline{G}$ the elements $\tilde{\mathrm{id}}_c\, \bar{g}\, \tilde{\mathrm{id}}_c^{-1}$ and $\bar{g}'$ from $\overline{G}'$ both project to $g \in G$ we must have $\bar{g}' = \tilde{\mathrm{id}}_c\, \bar{g}\, \tilde{\mathrm{id}}_c^{-1}$, for all $g \in G$. Rewrite as

$$\bar{g}'\, \tilde{\mathrm{id}}_c = \tilde{\mathrm{id}}_c\, \bar{g},$$

and let $\bar{g}(u_0, 0) = (gu_0, t_g)$ and $\bar{g}'(u_0, 0) = (gu_0, t_g')$. Then the left hand side maps the vertex $(u_0, 0)$ to $(gu_0, t_g' + g^\#(c))$, while the right hand side maps $(u_0, 0)$ to $(gu_0, t_g + c)$. Hence $t_g' + g^\#(c) = t_g + c$, and so

$$t_g - t_g' = \delta_c(g),$$

where $\delta_c \in \mathrm{Inn}(G, \Gamma)$ is an inner derivation. In particular, the above relation holds for $(t_1, t_2, \ldots, t_n)$ and $(t_1', t_2', \ldots, t_n')$ from Ω giving rise to $\overline{G}$ and $\overline{G}'$.

For the converse, let $(t_1, t_2, \ldots, t_n)$ and $(t_1', t_2', \ldots, t_n')$ from Ω give rise to $\overline{G}$ and $\overline{G}'$ such that $t_i - t_i' = \delta_c(g_i)$ for $i = 1, 2 \ldots, n$. Then we can work backwards to find that $\bar{g}_i' = \tilde{\mathrm{id}}_c\, \bar{g}_i\, \tilde{\mathrm{id}}_c^{-1}$ for all indices. Hence $\overline{G}$ and $\overline{G}'$ are conjugate subgroups. This completes the proof. □

Remark 5.3. Theorem 5.2 can be used to compute the first cohomology group $H^1(G, \Gamma)$, c.f. [23, Section 7.6]. Next, observe that each solution $(t_1, t_2, \ldots, t_n) \in \Omega$ extends uniquely to a function $t\colon G \to \Gamma$ satisfying condition (4.1), and that two such functions differ precisely by a derivation, $t' - t \in \mathrm{Der}(G, \Gamma)$. Thus, the set of functions $G \to \Gamma$ satisfying condition (4.1) forms a coset of $\mathrm{Der}(G, \Gamma)$ in the group of all functions $G \to \Gamma$ equipped with pointwise addition. Consequently, an alternative proof of the last statement in Theorem 5.2 can be given using the standard result which states that two derivations give rise to conjugate complements if and only if they differ by an inner derivation. □

Remark 5.4. Algorithm Evaluate requires some precomputations. First, we need to compute the vectors $\underline{\zeta_Q} \in \mathbb{Z}^{r,1}$, for some $Q : v \to u_0$ and all $v \in V(X)$, and consequently, the vectors representing voltages of fundamental walks at u_0. This can be done efficiently using breadth first search. During the search we also compute the vectors representing voltages of the mapped paths in order to obtain, upon completion of the search, the vectors $\underline{\zeta_{g_iQ}} \in \mathbb{Z}^{r,1}$ together with the vectors representing voltages of the mapped fundamental walks, for each g_i. Second, with these data in hand we then build the systems of linear equations over $\mathbb{Z}$ whose solutions give rise to the matrices $M_i \in \mathbb{Z}^{r,r}$ representing $g_i^\#$. □

Theorem 5.2 can also be used for testing whether a given group lifts as a direct product extension. In view of Corollary 4.5 we first check if the condition $\zeta_{gW} = \zeta_W$ holds for all $g \in S$ and all closed walks W from a basis of $H_1(X, \mathbb{Z})$. If true then G lifts, and we test whether the extension splits by solving the linear system (5.3) using algorithm Evaluate with all $g_i^\# = \mathrm{id}$. Algorithm Evaluate simplifies in that all matrices M_i are now equal to the identity matrix. Also, since the covering projection is abelian, recall that if some complement to $\mathrm{CT}_\wp$ is normal, then all complements are normal. We record this formally as Corollary 5.5.

Corollary 5.5. *With assumptions and notation above,* $\mathrm{id} \to \mathrm{CT}_\wp \to \tilde{G} \to G \to \mathrm{id}$ *is a direct product extension if and only if the following two conditions are satisfied:*

(i) $\zeta_{gW} = \zeta_W$ *holds for all* $g \in S$ *and all closed walks* W *from a basis of* $H_1(X, \mathbb{Z})$*;*

(ii) the (simplified) system of linear equations (5.3) has a solution in $\mathbb{Z}$*.*

Moreover, in this case the set of solutions of (5.3), reduced relative to the defining relations in Λ*, is in bijective correspondence with normal complements to* $\mathrm{CT}_\wp$ *within* $\tilde{G}$*.* □

5.2 Elementary abelian covers

One particular special case worth mentioning is that of $\mathrm{CT}_\wp$ being elementary abelian. In this case, Γ can be identified with the vector space over the corresponding prime field, and the automorphisms of Γ are then viewed as invertible linear transformations. More precisely, let $\Gamma = \mathbb{Z}_p^r$. The generating set $\{c_1, c_2, \ldots, c_r\}$ is now understood to be the standard generating set (of a vector space), and (5.1) can be viewed as a formula in this vector space. Consequently, instead of (5.3) we need to find solutions of (5.2) over $\mathbb{Z}_p$, which can be done using Gaussian elimination. This makes computation easier; in particular, we do not experience difficulties which might otherwise be present with computations over $\mathbb{Z}$ (like uncontrolled integer growth). An algorithm for testing whether the extension is split now immediately follows from the above discussion. It is formally encoded in algorithm IsSplit.

Algorithm: IsSplit

Input: Cayley voltage assignment $\zeta\colon D(X) \to \mathbb{Z}_p^d$ giving rise to connected cover, automorphism group $G = \langle g_1, g_2, \ldots, g_n \mid R_1, R_2, \ldots, R_m\rangle$ that lifts
Output: true, if the lifted group is split extension, false otherwise

set $\mathbf{A} \in \mathbb{Z}^{0,dn}$ to be the zero matrix and $\underline{b} \in \mathbb{Z}^{0,1}$ the zero vector;
take an arbitrary vertex u_0 in X;
compute set T of $|V(X)|$ vectors $\underline{\zeta_Q} \in \mathbb{Z}_p^{d,1}$ with $Q : v \to u_0$ for all $v \in V(X)$;
compute list $\mathcal{Z}$ of n sets each containing $|V(X)|$ vectors $\underline{\zeta_{g_iQ}} \in \mathbb{Z}_p^{d,1}$;
compute list $\mathcal{M}$ of n matrices $M_i \in \mathbb{Z}^{d,d}$ representing $g_i^{\#}$;
for $j \leftarrow 1$ **to** m **do**
  let $\mathbf{A}_j$ and $\underline{b}_j$ be the output of evaluating the relator R_j at $(u_0, 0)$ using the algorithm Evaluate;
  $\mathbf{A} \leftarrow \begin{bmatrix} \mathbf{A} \\ \mathbf{A}_j \end{bmatrix}$; $\underline{b} \leftarrow \begin{bmatrix} \underline{b} \\ \underline{b}_j \end{bmatrix}$;
if system $\mathbf{A}\,\mathbf{t} = -\underline{b}$ has a solution **then**
  return true
else
  return false

Theorem 5.6. *Let* $\wp = \wp_\zeta\colon \mathrm{Cov}(\zeta) \to X$ *be an elementary abelian regular covering projection of connected graphs arising from a Cayley voltage assignment* $\zeta\colon D(X) \to \mathbb{Z}_p^d$*. Further, let a given group of automorphisms* $G = \langle g_1, g_2, \ldots, g_n \mid R_1, R_2, \ldots, R_m\rangle$ *lift* *along $\wp$. Then the algorithm* IsSplit *tests whether the lifted group is a split extension of* $\mathrm{CT}_\wp$ *by* G *in*

$$\mathcal{O}(n|V(X)| + nd|D(X)| + d^3r + nd^2r + nd^3L + nd^3m^2)$$

steps using

$$\mathcal{O}(n|V(X)| + nd|D(X)| + nd^2m)$$

memory space, where r is the Betti number of X and $L = \sum_{j=1,2,\ldots,m} |R_j|$.

Proof. It remains to consider time and space complexity. The vectors $\underline{\zeta_{W_k}}$ representing the voltages of the fundamental walks W_k, $k = 1, 2, \ldots, r$, at u_0 together with the vectors $\underline{\zeta_{g_iW_k}}$ representing the voltages of the mapped fundamental walks g_iW_k, $i = 1, 2, \ldots, n$, as well as the vectors $\underline{\zeta_Q}$ and $\underline{\zeta_{g_iQ}}$ can be computed as described in Remark 5.4 using breadth first search at the cost of $\mathcal{O}(d)$ steps per edge; altogether this takes $\mathcal{O}(n|V(X)| + nd|D(X)|)$ steps. As for constructing the matrices $M_i \in \mathbb{Z}^{d,d}$ we first need to solve d systems of linear equations:

$$\begin{array}{rcl} x_{1,1}\,\underline{\zeta_{W_1}} + x_{1,2}\,\underline{\zeta_{W_2}} + \cdots + x_{1,r}\,\underline{\zeta_{W_r}} &=& e_1 \\ x_{2,1}\,\underline{\zeta_{W_1}} + x_{2,2}\,\underline{\zeta_{W_2}} + \cdots + x_{2,r}\,\underline{\zeta_{W_r}} &=& e_2 \\ &\vdots& \\ x_{d,1}\,\underline{\zeta_{W_1}} + x_{d,2}\,\underline{\zeta_{W_2}} + \cdots + x_{d,r}\,\underline{\zeta_{W_r}} &=& e_d, \end{array} \tag{5.4}$$

where e_i's are the standard basis vectors of $\mathbb{Z}_p^{d,1}$. Solving d systems using Gaussian elimination requires $\mathcal{O}(d^3r)$ steps. An arbitrary matrix M_i can then be computed in $\mathcal{O}(d^2r)$ steps; thus $\mathcal{O}(nd^2r)$ steps are required to compute n such matrices. Algorithm Evaluate takes $\mathcal{O}(nd^3|R_j|)$ steps for evaluating an arbitrary relator R_j. Hence all relators can be evaluated in $\mathcal{O}(nd^3L)$ steps, where $L = \sum_{j=1,2,\ldots,m} |R_j|$. It remains to solve the system $\mathbf{A}\,\underline{b} = -\mathbf{b}$ for $\mathbf{A} \in \mathbb{Z}_p^{dm,dn}$ and $\underline{b} \in \mathbb{Z}_p^{dn,1}$, which takes $\mathcal{O}(nd^3m^2)$ steps using Gaussian elimination. Hence the problem of testing whether the extension splits can be solved in $\mathcal{O}(n|V(X)| + nd|D(X)| + d^3r + nd^2r + nd^3L + nd^3m^2)$ steps.

Representing the graph X using adjacency list takes $\mathcal{O}(|V(X)|+|D(X)|)$ space, while representing a vector in $\mathbb{Z}_p^{d,1}$ takes $\mathcal{O}(d)$ space. Therefore the representation of the Cayley voltage assignment ζ takes $\mathcal{O}(|V(X)|+d|D(X)|)$ space. As for the representation of automorphisms as permutations, this takes $\mathcal{O}(n|D(X)|)$ space. During breadth first search we also need $\mathcal{O}(n|V(X)|)$ space to store the mapped vertices, and $\mathcal{O}(nd|D(X)|)$ additional space to store the voltages of the mapped walks. It takes $\mathcal{O}(nd^2)$ space to store all the matrices M_i, while storing the matrix $\mathbf{A} \in \mathbb{Z}_p^{dm,dn}$ takes $\mathcal{O}(nd^2m)$ space. Putting together, the space complexity is $\mathcal{O}(n|D(X)| + nd|D(X)| + nd^2m)$. □

Example 5.7. Let X be the 3-dipole with vertices 1 and 2 and three parallel links from 1 to 2 defined by the darts a, b and c. The voltage assignment $\zeta_a = \zeta_{a^{-1}} = (0,0)$, $\zeta_b = \zeta_{b^{-1}} = (1,0)$, $\zeta_c = \zeta_{c^{-1}} = (0,1)$ taking values in the elementary abelian group $\mathbb{Z}_2 \times \mathbb{Z}_2$ gives rise to a connected covering graph $\tilde{X}$ isomorphic to the 3-cube graph. Consider the group $G = \langle \sigma, \tau \mid \tau^2 = \sigma^3 = \tau\sigma\tau\sigma^2 = 1 \rangle$ acting as a subgroup of $\mathrm{Aut}\,X$, where $\sigma = (abc)(a^{-1}b^{-1}c^{-1})$ and $\tau = (aa^{-1})(bb^{-1})(cc^{-1})$. By computation, G lifts along $\wp_\zeta$. We now test whether $\wp_\zeta$ is split-admissible for the group G.

Choosing $u_0 = 1$ as the base vertex, let us work through the computation on the first relator τ^2. Observe that $M_\tau = \left[\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right]$. Set $\mathbf{E}_\tau = \left[\begin{smallmatrix}0&0&1&0\\0&0&0&1\end{smallmatrix}\right]$. Initially we have $\mathbf{A}_1 = \left[\begin{smallmatrix}0&0&0&0\\0&0&0&0\end{smallmatrix}\right]$, $\underline{b}_1 = \left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$, and $v = 1$. Scanning the relator from right to left we start with the generator τ. We multiply $\mathbf{A}_1$ on the left by M_τ, and then add $\mathbf{E}_\tau$ to get $\mathbf{A}_1 = \left[\begin{smallmatrix}0&0&1&0\\0&0&0&1\end{smallmatrix}\right]$. For the walk $W = 1a2b^{-1}1$ we have $\underline{\mu_W} = \underline{\mu_{\tau W}} = \left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$. So we add $\underline{\mu_W}$ to $\underline{b}_1$, multiply the result on the left by M_τ, and subtract $\underline{\mu_{\tau W}}$ to obtain $\underline{b}_1 = \left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$. Further, mapping the vertex v by τ we get $v = 2$. Moving left we scan the generator τ again. Multiplying $\mathbf{A}_1$ on the left by M_τ and adding $\mathbf{E}_\tau$ gives $\mathbf{A}_1 = \left[\begin{smallmatrix}0&0&0&0\\0&0&0&0\end{smallmatrix}\right]$. For the walk $W = 2c^{-1}1$ we have $\underline{\mu_W} = \underline{\mu_{\tau W}} = \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$. We then add $\underline{\mu_W}$ to $\underline{b}_1$, multiply the result on the left by M_τ, and subtract $\underline{\mu_{\tau W}}$ to get $\underline{b}_1 = \left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$.

Similarly, the computation on the second relator σ^3 gives $\mathbf{A}_2 = \left[\begin{smallmatrix}0&0&0&0\\0&0&0&0\end{smallmatrix}\right]$ and $\underline{b}_2 = \left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$, while the computation on the third relator $\tau\sigma\tau\sigma^2$ results in $\mathbf{A}_3 = \left[\begin{smallmatrix}0&0&1&1\\0&0&1&0\end{smallmatrix}\right]$ and $\underline{b}_3 = \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$. Putting together we obtain the following system over $\mathbb{Z}_2$

$$\begin{bmatrix}0&0&0&0\\0&0&0&0\\0&0&0&0\\0&0&0&0\\0&0&1&1\\0&0&1&0\end{bmatrix}\begin{bmatrix}x\\y\\u\\v\end{bmatrix} = \begin{bmatrix}0\\0\\0\\0\\0\\1\end{bmatrix},$$

which is clearly consistent. Thus the projection $\wp_\zeta$ is G-split-admissible. □

5.3 Solvable covers

The elementary abelian version of Theorem 5.2 can be used to decide whether G lifts as a split extension of $\mathrm{CT}_\wp$ whenever $\mathrm{CT}_\wp$ is solvable.

First recall, c.f. [35, 51], that if $q\colon Z \to X$ is a regular covering projection of connected graphs, and $q = r\,s$ where $r\colon Y \to X$ and $s\colon Z \to Y$ are regular covering projections with CT_s a characteristic subgroup of CT_q, then q is admissible for a group of automorphisms $G \le \mathrm{Aut}\,X$ if and only if G lifts along r and its lift lifts along s (in which case this lift is the lift of G along q).

The following lemma (c.f. [5, Theorem 4.2]) shows that testing whether the projection $q\colon Z \to X$ is split-admissible can be reduced to testing whether the projections $r\colon Y \to X$ and $s\colon Z \to Y$ are split-admissible. We omit the obvious proof.

Lemma 5.8. *Let $q\colon Z \to X$ be a regular covering projection of connected graphs, and let $q = r\,s$ where $r\colon Y \to X$ and $s\colon Z \to Y$ are regular covering projections with CT_s a characteristic subgroup of CT_q. Suppose that q is admissible for a group of automorphisms $G \le \mathrm{Aut}\,X$. Then the following statements are equivalent.*

(i) *The projection q is split-admissible for G.*

(ii) *The projection r is split-admissible for G, and s is split-admissible for some complement to CT_r within the G-lift along r.* □

Remark 5.9. Denote by $\tilde{G}$ the lift of G along $q\colon Z \to X$ and by H the lift of G along $r\colon Y \to X$. Observe that the projection $s\colon Z \to Y$ as in Lemma 5.8(ii) should be checked, at least in principle, relative to all complements of CT_r within H (in particular, this requires the construction of all such complements). However, if K is a complement, then any subgroup conjugate to K is also a complement; and if K lifts along $s\colon Z \to Y$ as a split extension, then any of its conjugate complements also lifts as a split extension. Therefore, when applying Lemma 5.8(ii) we only need to consider representatives of conjugacy classes of complements within H. A method for constructing such representatives is described in [5, 23]. □

Coming back to the case when $\mathrm{CT}_\wp$ is solvable, we first find a series of characteristic subgroups $\mathrm{CT}_\wp = K_0 > K_1 > \ldots > K_n = \mathrm{id}$ with elementary abelian factors K_{j-1}/K_j. The method is known, see [23, Chapter 8]. The covering projection $\wp$ then decomposes as $\tilde{X} = X_n \stackrel{\wp_n}{\to} X_{n-1} \to \ldots \to X_1 \stackrel{\wp_1}{\to} X_0 = X$, where $\wp_j \colon X_j \to X_{j-1}$ is a regular elementary abelian covering projection with $\mathrm{CT}_{\wp_j}$ isomorphic to K_{j-1}/K_j. At each step we may then recursively apply Lemma 5.8. To this end, one has to explicitly construct (among other things) the voltage assignments that define the intermediate projections in the above decomposition. For more details we refer the reader to [44].

6 Concluding remarks

In order to evaluate the performance of the above method for testing whether a given solvable covering projection is split-admissible the second author has implemented it in MAGMA [4], as a part of a larger package for computing with graph covers, see [43], and [44] for a more detailed account on experimental results.

We further remark that in the case of solvable covers one can take an alternative approach that even does not require explicit reconstruction of the intermediate covering projections. It is enough to first compute the automorphisms $g^{\#_{u_0}}$ and the factor set $\mathcal{F}(g,h) = g^{\#_{u_0}}(\zeta_Q)\zeta_{gQ}^{-1}$ (partially as needed) in order to reconstruct the lifted group as a crossed product, and then consider the decomposition abstractly without reference to covers. This is discussed in [45].

Acknowledgement. The authors would like to thank the referee for detailed comments that helped us improve the presentation.

References

[1] D. Archdeacon, P. Gvozdjak and J. Širáň, Constructing and forbidding automorphisms in lifted maps, *Math. Slovaca* **47** (1997), 113–129.

[2] S. Bhoumik, T. Dobson and J. Morris, On the automorphism groups of almost all circulant graphs and digraphs, *Ars Math. Contemp.* **7** (2014), 499–518.

[3] N. L. Biggs, *Algebraic Graph Theory*, Cambridge Univ. Press, Cambridge, 1974.

[4] W. Bosma, J. Cannon and C. Playoust, The MAGMA algebra system I: The user language, *J. Symbolic Comput.* 24 (1997), 235–265.

[5] F. Celler, J. Neubser and C. R.B. Wright, Some remarks on the computation of complements and normalizers in solvable groups, *Acta Applicandae Mathematicae* **21** (1990), 57–76.

[6] M. D. E. Conder and P. Dobcsányi, Trivalent symmetric graphs on up to 768 vertices, *J. Combin. Math. Combin. Comput.* **40** (2002), 41–63.

[7] M. D. E. Conder and J. Ma, Arc-transitive abelian regular covers of cubic graphs, *J. Algebra* **387** (2013), 215–242.

[8] M. D. E. Conder, A. Malnič, D. Marušič and P. Potočnik, A census of cubic semisymmetric graphs on up to 768 vertices, *J. Algebraic Combin.* **23** (2006), 255–294.

[9] D. Ž. Djoković, Automorphisms of graphs and coverings, *J. Combin. Theory Ser. B* **16** (1974), 243–247.

[10] D. Ž. Djoković and G. L. Miller, Regular groups of automorphisms of cubic graphs, *J. Combin Theory* **29** (1980), 195–230.

[11] S. F. Du, J. H. Kwak and M. Y. Xu, 2-arc-transitive regular covers of complete graphs having the covering transformation group $\mathbb{Z}_p^3$, *J. Combin. Theory Ser. B* **93** (2005), 73-93.

[12] S. F. Du, A. Malnič and D. Marušič, Classification of 2-arc-transitive dihedrants, *J. Combin. Theory Ser. B* **98** (2008), 1349–1372.

[13] S. F. Du, D. Marušič and A. O. Waller, On 2-arc transitive covers of complete graphs, *J. Combin. Theory Ser. B* **74** (1998), 276–290.

[14] Y. Q. Feng and J. H. Kwak, Classifying cubic symmetric graphs of order $10p$ or $10p^2$, *Science China Ser. A: Math.* **49** (2006), 300-319.

[15] Y. Q. Feng, A. Malnič, D. Marušič and K. Kutnar, On 2-fold covers of graphs, *J. Combin. Theory Ser. B* **98** (2008), 324–341.

[16] M. Giudici, C. H. Li and C. E. Praeger, Analysing finite locally s-arc transitive graphs, *Trans. Amer. Math. Soc.* **356** (2004), 291–317.

[17] R. Gramlich, G. W. Hofmann and K. H. Neeb, Semi-edges, reflections and Coxeter groups, *Trans. Amer. Math. Soc.* **359** (2007), 3647–3668.

[18] J. L. Gross and S. R. Alpert, Branched coverings of graph imbeddings, *Bull. Amer. Math. Soc.* **79** (1973), 942–945.

[19] J. L. Gross, Voltage graphs, *Discrete Math.* **9** (1974), 239–246.

[20] J. L. Gross and T. W. Tucker, *Topological Graph Theory,* Wiley - Interscience, New York, 1987.

[21] M. Hofmeister, Isomorphisms and automorphisms of graph coverings, *Discrete Math.* **98** (1991), 175–183.

[22] M. Hofmeister, Graph covering projections arising from finite vector spaces over finite fields, *Discrete Math.* **143** (1995), 87–97.

[23] D. Holt, B. Eick and E. A. O'Brien, *Handbook of Computational Group Theory,* Chapman and Hall/CRC, Boca Raton London New York Washington D.C., 2005.

[24] A. Hujdurović, Quasi m-Cayley circulants, *Ars Math. Contemp.* **6** (2013), 147–154.

[25] H. Koike and I. Kovács, Isomorphic tetravalent cyclic Haar graphs, *Ars Math. Contemp.* **7** (2014), 215–235.

[26] I. Kovács, K. Kutnar, D. Marušič, Classification of edge-transitive rose window graphs, *J. Graph Theory* **65** (2010), 216–231.

[27] I. Kovács, B. Kuzman and A. Malnič, On non-normal arc transitive 4-valent dihedrants, *Acta Math. Sin. (Engl. ser.)* **26** (2010), 1485–1498.

[28] B. Kuzman, Arc-transitive elementary abelian covers of the complete graph K_5, *Linear Algebra Appl.* **433** (2010), 1909–1921.

[29] J. H. Kwak and J. Lee, Counting some finite-fold coverings of a graph, *Graphs Combin.* **8** (1992), 277–285.

[30] J. H. Kwak and J. M. Oh, Arc-transitive elementary abelian covers of the octahedron graph, *Linear Algebra Appl.* **429** (2008), 2180–2198.

[31] P. Lorimer, Vertex-transitive graphs: Symmetric graphs of prime valency, *J.$\tilde{G}$raph Theory* **8** (1984), 55–68.

[32] A. Malnič, Group actions, coverings and lifts of automorphisms, *Discrete Math.* **182** (1998), 203–218.

[33] A. Malnič, R. Nedela and M. Škoviera, Lifting graph automorphisms by voltage assignments, *European J. Combin.* **21** (2000), 927–947.

[34] A. Malnič, Action graphs and coverings, *Discrete Math.* **244** (2002), 299–322.

[35] A. Malnič, D. Marušič and P. Potočnik, Elementary abelian covers of graphs, *J. Alg. Combin.* **20** (2004), 71–96.

[36] A. Malnič, D. Marušič and P. Potočnik, On cubic graphs admitting an edge-transitive solvable group, *J. Algebraic Combin.* **20** (2004), 99–113.

[37] A. Malnič and P. Potočnik, Invariant subspaces, duality, and covers of the Petersen graph, *European J. Combin.* **27** (2006), 971–989.

[38] S. McLane, *Homology*, Springer-Verlag, Berlin Gottingen Heidelberg, 1963.

[39] A. Mednykh and R. Nedela, Enumeration of unrooted maps a given genus, *J. Combin. Theory Ser. B* **96** (2006), 706–729.

[40] P. Potočnik, Edge-colourings of cubic graphs admitting a solvable vertex-transitive group of automorphisms, *J. Combin. Theory Ser. B* **91** (2004), 289-300.

[41] P. Potočnik, M. Škoviera and R. Škrekovski, Nowhere-zero 3-flows in abelian Cayley graphs, *Discrete math.* **297** (2005), 119–127.

[42] R. Požar, Sectional split extensions arising from lifts of groups, *Ars Math. Contemp.* **6** (2013), 393–408.

[43] R. Požar, `http://osebje.famnit.upr.si/~ rok.pozar`

[44] R. Požar, Some computational aspects of solvable regular covers of graphs, *J. Symbolic Comput.* (2014), doi: 10.1016/j.jsc.2014.09.023

[45] R. Požar, Testing whether the lifted group splits, *Ars Math. Contemp.*, to appear.

[46] C. E. Praeger, Imprimitive symmetric graphs, *Ars Combin.* **19** (1985), 149–163.

[47] C. E. Praeger, An O'Nan-Scott Theorem for finite quasiprimitive permutation groups, and an application to 2-arc transitive graphs, *J. London Math. Soc.(2)* **47** (1993), 227–239.

[48] S. Rees, L. H. Soicher, An algorithmic approach to fundamental groups and covers of combinatorial cell complexes, *J. Symbolic Comput.* **29** (2000), 59–77.

[49] J. Širáň, Coverings of graphs and maps, orthogonality, and eigenvectors, *J. Algebraic Combin.* **14** (2001), 57–72.

[50] M. Škoviera, A contribution to the theory of voltage graphs, *Discrete Math.* **61** (1986), 281–292.

[51] A. Venkatesh, Covers in imprimitevely symmetric graphs, Honours dissertation, Department of Mathematics and Statistics, University of Western Australia, 1997.

[52] C. Q. Wang and Y. H. Hao, Edge-transitive regular $\mathbb{Z}_n$-covers of the Heawood graph, *Discrete Math.* **310** (2010), 1752–1758.

[53] L. Wang, S. F. Du and X. Li, A class of semisymmetric graphs, *Ars Math. Contemp.* **7** (2014), 40–53.

[54] A. T. White, *Graphs, Groups, and Surfaces,* North-Holland, Amsterdam, 1984 (rev. ed.).

[55] J. X. Zhou and Y. Q. Feng, Semisymmetric elementary abelian covers of the Heawood graph, *Discrete Math.* **310** (2010), 3658–3662.

Also available at http://amc-journal.eu
ISSN 1855-3966 (printed edn.), ISSN 1855-3974 (electronic edn.)
ARS MATHEMATICA CONTEMPORANEA 10 (2015) 135–167

Unilateral and equitransitive tilings by squares of four sizes

Casey Mann
University of Washington – Bothell, School of STEM, 18115 Campus Way N.E., Bothell, WA 98011, USA

Joseph DiNatale
Armstrong Atlantic State University, USA

Emily Peirce
Baylor University, USA

Ellen Vitercik
Columbia University, USA

Received 3 April 2014, accepted 15 July 2014, published online 18 August 2015

Abstract

D. Schattschneider proved that there are exactly eight unilateral and equitransitive tilings of the plane by squares of three distinct sizes. This article extends Schattschneider's methods to determine a classification of all such tilings by squares of four different sizes. It is determined that there are exactly 39 unilateral and equitransitive tilings by squares of four different sizes.

Keywords: Tilings, equitransitive, unilateral, squares.

Math. Subj. Class.: 05B45, 52C20, 54E15, 05C15

1 Introduction

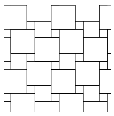
A two-dimensional *tiling*, $\mathscr{T}$, is a countable collection of closed topological disks $\{T_i\}$, called *tiles*, such that the interiors of the T_i are pairwise disjoint and the union of the T_i is the Euclidean plane. A *symmetry* of $\mathscr{T}$ is any planar isometry that maps every tile of $\mathscr{T}$ onto a tile of $\mathscr{T}$. Two tiles T_1 and T_2 are *equivalent* if there exists a symmetry of $\mathscr{T}$ that

E-mail addresses: cmann@uwb.edu (Casey Mann), jd4732@stu.armstrong.edu (Joseph DiNatale), EmilyPeirce@baylor.edu (Emily Peirce), emv2126@columbia.edu (Ellen Vitercik)

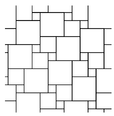
maps T_1 onto T_2. The collection of all tiles of $\mathscr{T}$ that are equivalent to T_1 is called the *transitivity class* of T_1. $\mathscr{T}$ is *equitransitive* if each set of mutually congruent tiles forms one transitivity class.

This article will concern only tilings of the plane by squares of a few different sizes. A connected segment formed by the intersection of two squares of $\mathscr{T}$ will be called an *edge* of $\mathscr{T}$, and the endpoints of the edges are called *vertices* of $\mathscr{T}$. $\mathscr{T}$ is *unilateral* if each edge of the tiling is a side of at most one tile, meaning that if two congruent tiles meet along an edge, they are never incident along the full length of that edge. The acronym *UETn* will refer to a unilateral and equitransitive tiling by squares of n distinct sizes.

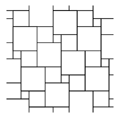
A classification of all UET3 tilings is given in [3]. There are only eight UET3 tilings, shown in Figure 1. Because the classification of UET4 tilings is based on the methology of [3], it will be helpful to outline those methods here. First, some notation and terminology must be introduced.



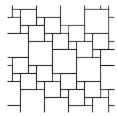
(a) $(c.c.c.b, c.a.c.c.a.c, a.c.a.c.b.a.c.b)$



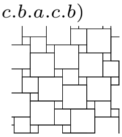
(b) $(c.c.b.b, c.a.b.a.c.c, b.c.b.a.c.b.a.c)$



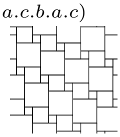
(c) $(c.b.c.b, c.c.a.c.c, b.c.b.c.b.a.b.c)$



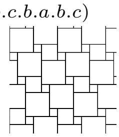
(d) $(c.b.b.b, a.c.c.a.b.a.b, a.b.c.b.a.b.c.b)$



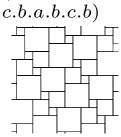
(e) $(c.c.c.b, a.c.c.a.c.c, a.b.c.a.b.c.a.c)$



(f) $(c.b.c.b, a.c.c.a.c.c, a.b.c.b.a.b.c.b)$



(g) $(c.c.c.b, a.c.c.a.c.c, a.c.a.c.a.b.c.b.a)$



(h) $(c.c.b.b, a.c.c.c.a.b, a.b.c.b.c.b.a.c)$

Figure 1: The eight UET3 tilings classified in [3].

Let $\mathscr{T}$ be a UET4 tiling of squares with side lengths $a < b < c < d$. The *skeleton* of $\mathscr{T}$ is the union of all of the edges of the tiling $\mathscr{T}$. A *vortex* is a tile $T \in \mathcal{T}$ for which each edge of the tile is extendable within the skeleton of $\mathscr{T}$ in exactly one direction, given an orientation of T, as in Figure 2.

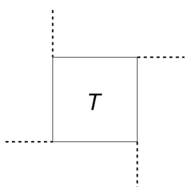


Figure 2: A vortex tile

A *corona* of a tile T in a tiling $\mathscr{T}$ consists of all tiles in $\mathscr{T}$ whose intersection with T is nonempty. The *corona signature* of T is an ordered list of the sizes of the tiles in T's corona. In a UET4 tiling the coronas of any two congruent copies of T must be congruent due to equitransitivity, so the corona signature of T unambiguously describes the corona of any tile in $\mathscr{T}$ that is congruent to T. A sample d corona (i.e. a corona of a d tile) and its corresponding corona signature are given in Figure 3. The corona signatures of the eight

UET3 tilings shown in Figure 1 are given as a triplet (a corona signature, b corona signature, c corona signature). Cyclic permutations of a signature, as well as cyclic permutations of a signature read in reverse, are considered equivalent to the original signature.

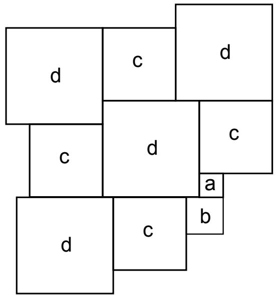


Figure 3: This d corona has signature $c.d.c.a.b.c.d.c.d$.

1.1 Schattschneider's Method

The method of classification used in [3] to find all UET3 tilings can be roughly described as follows:

1. Determine all extendable a, b, and c coronas.
2. Determine which 3-tuples of extendable a, b, and c coronas are compatible in terms of their corona signatures.
3. Determine which 3-tuples of compatible a, b, and c corona signatures give rise to tilings (and how many).

A similar process will be followed in this article (with the addition of finding all extendable d corona signatures). It is to be expected that the scope of the UET4 classification problem is broader in size than the UET3 problem; as a result, the problem is solved through two major cases. These are the cases where:

1. a and b are adjacent.
2. a and b are not adjacent.

Two tiles are *adjacent* if their intersection is an edge of the tiling. Sections 2 - 4 concern the case when a tiles and b tiles are adjacent. While the case where a tiles and b tiles are not adjacent employs themes established these sections, the differences between these cases are sufficient to require a separate analysis; this is done in Section 5.

2 UET4 tilings in which a tiles and b tiles are adjacent

The bulk of the work done in classifying all UET4 tilings is enumerating all possible a, b, c, and d coronas. This job is made manageable by first establishing some necessary equations relating a, b, c, and d. These equations are established in Subsection 2.1. After establishing a finite set of possible equations relating a, b, c, and d, the coronas corresponding to these equations are found; this is described in Section 3. Finally, once a set of coronas corresponding to an equation or equations is established, the process for constructing the possible tilings is described in Section 4.

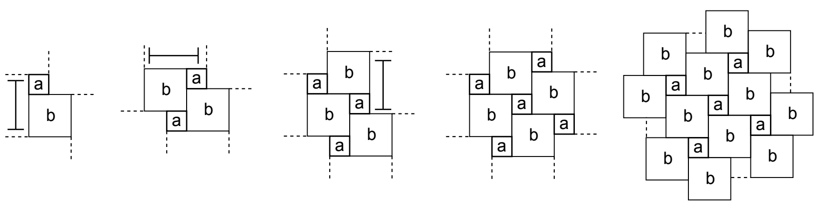
One fact that will be used throughout the article comes from [2], and can be stated as follows.

Lemma 2.1. *Let $\mathscr{T}$ be a UET4 tiling of squares with side lengths $a < b < c < d$. Then all a and b tiles of $\mathscr{T}$ are vortices.*

2.1 Equations relating *a*, *b*, *c*, and *d*

Lemma 2.2. *Let $\mathscr{T}$ be a UET4 tiling in which a and b are adjacent. Then $a + b = c$ or $a + b = d$.*

Proof. Begin by examining an a corona. Because a and b tiles are adjacent vortices, these two tiles must meet at a corner as shown in Figure 4a. The dashed lines depict the necessary skeletal extension in $\mathscr{T}$ required by the vortex condition on the a and b tiles. It is clear that a tile or group of tiles must fill the length indicated in Figure 4a exactly in order for these vortex conditions to hold.



(a) (b) (c) (d) (e)

Figure 4

Suppose $a+b \neq c$ and $a+b \neq d$. Neither a c tile nor a d tile can fill the space indicated exactly, so some combination of a and b tiles must be used instead. In fact, the vortex condition requires that exactly two such tiles be used, and unilaterality implies that exactly one a tile and one b tile must be used. This yields the arrangement shown in Figure 4b. The length indicated in 4b brings up the same issue, and following the same logic it is seen that the arrangement in Figure 4c is the only valid arrangement for this space. The same is true for the length indicated in Figure 4c, yielding the full a corona found in Figure 4d. Having now completed an a corona, equitransitivity tells us that every a corona in $\mathscr{T}$ must be identical to this, which generates the patch seen in Figure 4e. The only possible unilateral and equitransitive tiling that this patch admits contains only a and b tiles and is therefore UET2. □

This gives rise to two subcases within the case of a and b being adjacent, namely that where $a + b = c$ and that where $a + b = d$. These two cases are considered in turn.

2.1.1 $a + b = c$

The following two subcases of that when a and b are adjacent and the equation $a + b = c$ is satisfied are considered separately:

1. The d tile is not a vortex.
2. The d is a vortex.

The d tile is not a vortex:

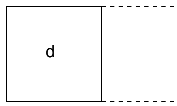


Figure 5: The extension of a non-vortex d tile.

If d is not a vortex, then is must have a pair of parallel edges that extend into the skeleton of $\mathscr{T}$ as in Figure 5. There must be some combination of a, b, and c tiles which fit perfectly between the dashed lines in Figure 5. Since a and b are vortices, they must share a corner with the d tile and therefore one of their edges must be contained in a dashed line. There are exactly five possible ways to fill the region, all of which are shown in Figure 6.

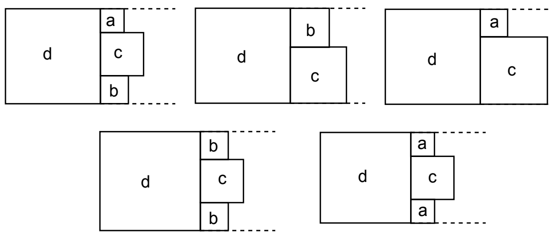


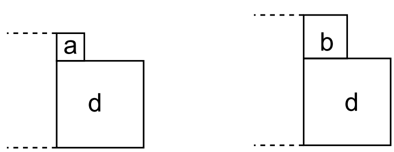
Figure 6: All possible ways of filling the region between the dashed lines.

This then gives us exactly five possible relationships for d if it is not a vortex:

1. $d = a + b + c = 2a + 2b$
2. $d = b + c = a + 2b$
3. $d = a + c = 2a + b$
4. $d = 2b + c = 3b + a$
5. $d = 2a + c = 3a + b$

The tile d is a vortex:

If d is a vortex, then the d corona must contain either an a or a b, as explained below; furthermore, there must be an a or b tile that shares a corner with the d tile to satisfy the pertinent vortex conditions, as depicted in Figure 7.



(a) (b)

Figure 7

If this was not the case, then the tiling would be UET2. To see this, notice that since d is a vortex, there must be tiles which line up exactly with the doted lines in Figure 8a. If there are no a or b tiles in the corona of the d tile, these tiles must be c tiles, as in Figure 8b. Finally, the rest of the corona must be made up by d tiles, as in Figure 8c. This patch can only be extended to a UET2 tiling. Therefore, there must be at least one a or b tile which shares a corner with the d tile.

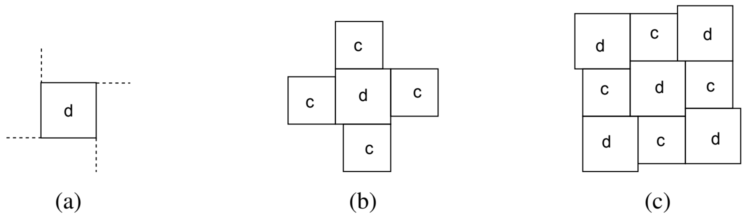


Figure 8: A UET2 corona.

Since a, b, and d are all vortices in this subcase, there must be a combination of tiles that fits perfectly between the dashed edges indicating edge extentions into the skeleton of the tiling in Figure 7. No more than three tiles may fit in this space, for if there were four or more, then the two or more tiles sandwiched in the middle would have to be non-vortices. However, only c can be a non-vortex, and two c tiles cannot meet edge-to-edge by the unilateral condition.

Only specific configurations of tiles can fit between the dashed lines. By examining these configurations and using simple algebra, it is easy to enumerate the possible relationships between d and the smaller square sizes that would allow for a tiling. This analysis is summarized in Table 1.

Of course, it is possible for a d tile to share a corner with a a or b tile even if d is not a vortex, which is why some d relations are repeated from the case where d tiles were not vortices. In this case, however, only those d relationships for which the tile is *necessarily* a vortex are considered, so any repeated relationships are disregarded, resulting in the following complete list of d relationships when $c = a + b$:

1. $d = a + b + c = 2a + 2b$
2. $d = b + c = a + 2b$
3. $d = a + c = 2a + b$
4. $d = 2b + c = 3b + a$
5. $d = 2a + c = 3a + b$
6. $d = 2b$
7. $d = 3a$
8. $d = 3b$

2.1.2 Aside: $a : b$ side length proportions

When $a + b = c$ and a and b are adjacent, there are certain tile configurations that are possible only when the size of b is specifically related to the size of a. By examining

All possible combinations of two or three tiles	Could they fit between the dashed lines in Figure 7a?	If yes, what d relations are required for the configuration to fit perfectly?	Could they fit between the dashed lines in Figure 7b?	If yes, what d relations are required for the configuration to fit perfectly?
$a+b$	No		No	
$a+c$	No		No	
$a+d$	Yes	$a+d=a+d$ $\rightarrow$ No new information	No	
$b+c=2b+a$	Yes	$a+d=a+2b$ $\rightarrow d=2b$	No	
$b+d$	No		Yes	$b+d=b+d$ $\rightarrow$ No new information
$c+d$	No		No	
$a+c+a=3a+2b$	No		Yes	$b+d=3a+b$ $\rightarrow d=3a$
$a+c+b=2a+2b$	Yes	$a+d=2a+2b$ $\rightarrow d=a+2b$	Yes	$b+d=2a+2b$ $\rightarrow d=2a+b$
$a+c+d$	No		No	
$b+c+b=a+3b$	Yes	$a+d=a+3b$ $\rightarrow d=3b$	No	
$b+c+d$	No		No	

Table 1

these specific configurations for each case, certain $a:b$ ratios are determined that must be considered; such ratios are determined when a set of tiles must fit perfectly between the extended edges of two vortices. These specific tile configurations are shown in Figure 9 and Figure 10. Tables showing the arithmetic used to find the $a:b$ ratios are provided as well. In the first row of both tables, the eight d relations previously generated for this case ($a+b=c$ and a and b are adjacent) are considered. In the leftmost column, the configurations as well as the general proportions they necessitate are listed. For example, in Figure 9a, a b tile and a c tile fit perfectly above two a's and a c. For this configuration to occur, $b+c=2a+c$. Therefore, $b=2a$.

It should be noted that the subcases within the case where $a+b=c$, each of which is technically given by a different b side length, are not considered within the actual corona construction process as separate cases. Instead, the reader should bear them in mind as corona construction begins within the appropriate specified subcase. It should additionally be noted that these $a:b$ side relations are only pertinent when $a+b=c$. They are not considered in the case where $a+b=d$, which follows.

2.1.3 $a+b=d$

Lemma 2.3. *If $a+b=d$, then c must be a vortex.*

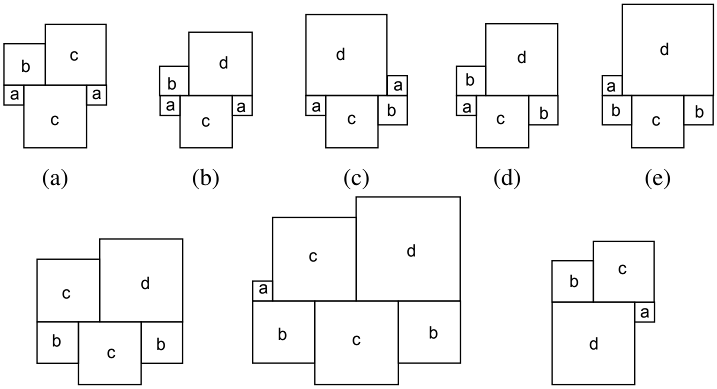
Proof. Suppose the c tile is not a vortex. Then two edges of any c tile must extend into the

	$d = 2a + b$	$d = a + 2b$	$d = 3a + b$	$d = 2a + 2b$	$d = a + 3b$	$d = 2b$	$d = 3b$	$d = 3a$
Tile Configurations and Associated General Proportions								
Figure 9a: $2a + c = c + b$ $\rightarrow b = 2a$	$\rightarrow b = 2a$	$\rightarrow b = 2a$	$\rightarrow b = 2a$	$\rightarrow b = 2a$	$\rightarrow b = 2a$	$\rightarrow b = 2a$	$\rightarrow b = 2a$	$\rightarrow b = 2a$
Figure 9b: $2a+c = 3a+b = b + d$ $\rightarrow d = 3a$	$3a = 2a + b$ $\rightarrow$ not possible	$3a = a + 2b$ $\rightarrow$ not possible	$3a = 3a + b$ $\rightarrow$ not possible	$3a = 2a+2b$ $\rightarrow$ not possible	$3a = a + 3b$ $\rightarrow$ not possible	$3a = 2b$ $\rightarrow$ not possible	$3a = 3b$ $\rightarrow$ not possible	$3a = 3b$ $\rightarrow$ not possible
Figure 9c: $a + b + c = 2a + 2b = a + d$ $\rightarrow d = a + 2b$	$a + 2b = 2a + b$ $\rightarrow$ not possible	$a+2b = a+ 2b$ $\rightarrow$ not possible	$a + 2b = 3a + b$ $\rightarrow b = 2a$	$a + 2b = 2a + 2b$ $\rightarrow$ not possible	$a+2b = a+ 3b$ $\rightarrow$ not possible	$a + 2b = 2b$ $\rightarrow$ not possible	$a + 2b = 2b$ $\rightarrow$ not possible	$a + 2b = 3b$ $\rightarrow$ not possible
Figure 9d: $a + b + c = 2a + 2b = b + d$ $\rightarrow d = 2a + b$	$2a + b = 2a + b$ $\rightarrow$ Not possible	$2a+b = a+ 2b$ $\rightarrow$ Not possible	$2a + b = 3a + b$ $\rightarrow$ Not possible	$2a + b = 2a + 2b$ $\rightarrow$ Not possible	$2a+b = a+ 3b$ $\rightarrow$ Not possible	$2a + b = 2b$ $\rightarrow b = 2a$	$2a + b = 3b$ $\rightarrow$ Not possible	$2a + b = 3a$ $\rightarrow$ Not possible
Figure 9e: $2b+c = a+3b = a + d$ $\rightarrow 3b = d$	$3b = 2a + b$ $\rightarrow$ Not possible	$3b = a + 2b$ $\rightarrow$ Not possible	$3b = 3a + b$ $\rightarrow b = \frac{3}{2}a$	$3b = 2a+2b$ $\rightarrow b = 2a$	$3b = a + 3b$ $\rightarrow$ Not possible	$3b = 2b$ $\rightarrow$ Not possible	$3b = 3b$ $\rightarrow$ Not possible	$3b = 3a$ $\rightarrow$ Not possible
Figure 9f: $2b + c = c + d$ $\rightarrow 2b = d$	$2b = 2a + b$ $\rightarrow b = 2a$	$2b = a + 2b$ $\rightarrow$ Not possible	$2b = 3a + b$ $\rightarrow b = 3a$	$2b = 2a+2b$ $\rightarrow$ Not possible	$2b = a + 3b$ $\rightarrow$ Not possible	$2b = 2b$ $\rightarrow$ Not possible	$2b = 3b$ $\rightarrow$ Not possible	$2b = 3a$ $\rightarrow b = \frac{3}{2}a$
Figure 9g: $2b+c = a+c+d$ $\rightarrow 2b = a + d$	$2b = 3a + b$ $\rightarrow b = 3a$	$2b = 2a+2b$ $\rightarrow$ Not possible	$2b = 4a + b$ $\rightarrow b = 4a$	$2b = 3a+2b$ $\rightarrow$ Not possible	$2b = 2a+3b$ $\rightarrow$ Not possible	$2b = a + 2b$ $\rightarrow$ Not possible	$2b = a + 3b$ $\rightarrow$ Not possible	$2b = 4a$ $\rightarrow b = 2a$
Figure 18b: $b + c = a + 2b = a + d$ $\rightarrow 2b = d$	$\rightarrow b = 2a$	$2b = a + 2b$ $\rightarrow$ Not possible	$\rightarrow b = 3a$	$2b = 2a+2b$ $\rightarrow$ Not possible	$2b = a + 3b$ $\rightarrow$ Not possible	$2b = 2b$	$2b = 3b$ $\rightarrow$ Not possible	$\rightarrow b = \frac{3}{2}a$

Table 2

Tile Configurations and Associated General Proportions	$d=2a+b$	$d=a+2b$	$d=3a+b$	$d=2a+2b$	$d=a+3b$	$d=2b$	$d=3b$	$d=3a$
Figure 10a: $2a+d=c+b=a+2b$ $\rightarrow a+d=2b$	$2b=3a+b$ $\rightarrow b=3a$	$2b=2a+2b$ $\rightarrow$ Not possible	$2b=4a+b$ $\rightarrow b=4a$	$2b=3a+2b$ $\rightarrow$ Not possible	$2b=2a+3b$ $\rightarrow$ Not possible	$2b=a+2b$ $\rightarrow$ Not possible	$2b=a+3b$ $\rightarrow$ Not possible	$2b=4a$ $\rightarrow b=2a$
Figure 10b: $2a+d=b+c+b=a+3b$ $\rightarrow a+d=3b$	$3b=3a+b$ $\rightarrow b=\frac{3}{2}a$	$3b=2a+2b$ $\rightarrow b=2a$	$3b=4a+b$ $\rightarrow b=2a$	$3b=3a+2b$ $\rightarrow b=3a$	$3b=2a+3b$ $\rightarrow$ Not possible	$3b=a+2b$ $\rightarrow$ Not possible	$3b=a+3b$ $\rightarrow$ Not possible	$3b=4a$ $\rightarrow b=\frac{4}{3}a$
Figure 10c: $2a+d=b+d$ $\rightarrow b=2a$	$\rightarrow b=2a$	$\rightarrow b=2a$	$\rightarrow b=2a$	$\rightarrow b=2a$	$\rightarrow b=2a$	$\rightarrow b=2a$	$\rightarrow b=2a$	$\rightarrow b=2a$
Figure 20b $a+b+d=c+d$ $\rightarrow a+b=c$ $\rightarrow$ No new information								
Figure 9e: $2b+d=a+c+d$ $2b=2a+b$ $\rightarrow b=2a$	$\rightarrow b=2a$	$\rightarrow b=2a$	$\rightarrow b=2a$	$\rightarrow b=2a$	$\rightarrow b=2a$	$\rightarrow b=2a$	$\rightarrow b=2a$	$\rightarrow b=2a$
Figure 10f: $2b+d=2a+c+d$ $2b=3a+b$ $\rightarrow b=3a$	$\rightarrow b=3a$	$\rightarrow b=3a$	$\rightarrow b=3a$	$\rightarrow b=3a$	$\rightarrow b=3a$	$\rightarrow b=3a$	$\rightarrow b=3a$	$\rightarrow b=3a$

Table 3



(a) (b) (c) (d) (e)

(f) (g) (h)

Figure 9

skeleton of the UET4 tiling $\mathscr{T}$ as in Figure 11:

It is apparent then that some tile or combination of tiles must fit exactly between the dashed lines shown above. A d tile is clearly to large to fit between these lines, and a c tile cannot be placed there by unilaterality. Since $a < b < c$, then some combination of tiles a and b must fit between these lines, and since both a and b must be vortices, then in fact only two tiles may fit between these dashed lines. Because $\mathscr{T}$ is unilateral, it is seen that one a tile and one b tile must fit exactly between the dashed lines in Figure 11. However, this implies that $d = a + b = c < d$, a contradiction. Therefore, there is no tile or combination of tiles that can fit exactly between the dashed lines above, so c must be a vortex. □

Next, enumerate possible ways to express c in terms of a, b, and d within the $a + b = d$ case. Note that $c < d = a + b$, and observe that when the c tile is surrounded d tiles (as in Figure 12 below), no further specifications as to values of c can be made.

Setting this special case aside momentarily, continue, using the fact that c must be a vortex, to find all possible relationships for c based on implications that arise through each of the three cases found in Figure 13:

Note that in the cases illustrated in 13a and 13b, the skeleton of the tiling $\mathscr{T}$ must extend along the dashed lines by virtue of a, b, and c all being vortices.

Begin with a statement implying the impossibility of the existence of the partial c corona in Figure 13a in a UET4 tiling.

Lemma 2.4. *If $a + b = d$ in a UET4 tiling, then each c corona will not contain an a.*

Proof. Let $\mathscr{T}$ be a UET4 tiling such that $a + b = d$ and suppose that the c tile's corona contains at least one a tile. Because both a and c are vortices, they must meet at one of their corners as shown in Figure 13a above; the dashed lines show the necessary skeleton extension of $\mathscr{T}$ also required by this vortex condition. Next, determine which tile or combination of tiles can fit exactly between the dashed lines that extend toward the left from the union of the left edges of a and c. Were one tile to fill this space, it would have to be a tile d, which would imply that $d = a + c > a + b$, a contradiction. Hence more than one tile must fill this space. Were three tiles to fill this space, then the middle tile in the

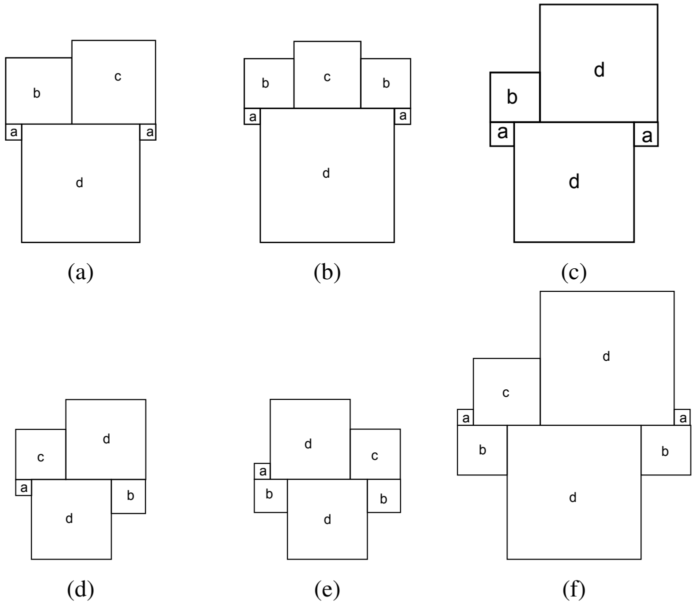


Figure 10

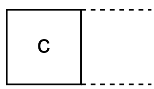


Figure 11: Necessary configuration when c is not a vortex

group must be a non-vortex tile; because d is the only non-vortex tile in this case, one of the three tiles must be a d. Then regardless of the other two tiles chosen, the sum of their side lengths will always exceed the length $a + c$ between the dashed lines. Hence three tiles cannot fill this space exactly; it is obvious that four or more tiles similarly cannot fill the space appropriately. This leaves the case where two tiles exactly fill the space between these dashed lines. Then all possible combinations of two distinct tiles are listed as follows: a and b; a and c; a and d; b and c; b and d; c and d. Of these combinations, the only one that covers the length $a + c$ exactly is the combination a and c. Therefore these two tiles must be placed along the left edges of a and c from Figure 13a, and this arrangement is shown below in Figure 14a along with the necessary skeleton extension required by the

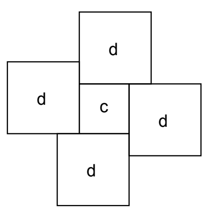
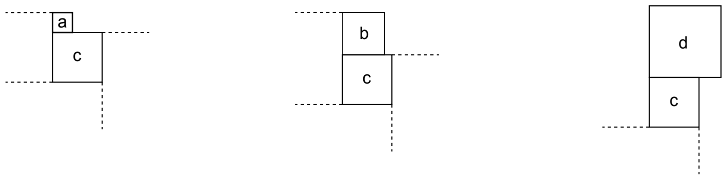


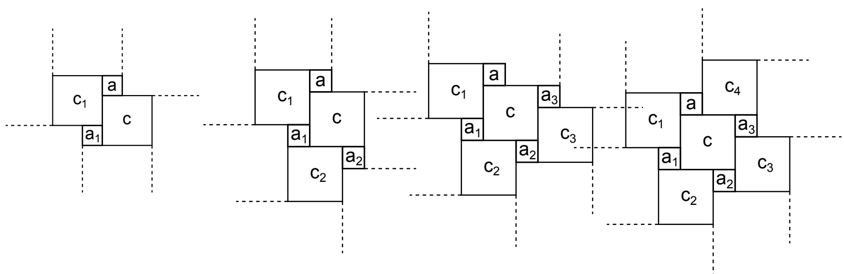
Figure 12: c surrounded by d's



(a) (b) (c)

Figure 13

vortex conditions of a and c. The same issue arises again: a tile or group of tiles must fit between the dashed lines that extend downward from the union of the bottom edges of tiles a_1 and c, and by the argument above, only tiles a and c can fill this space exactly. This logic is repeated again in Figure 14b to arrive at the partial c corona in Figure 14c, and it is clear by our assumptions that only a tile c can be placed along the top edge of tile a_3 to complete the corona, which is shown in Figure 14d. By equitransitivity, the tiling that this patch generates is in fact UET2. □



(a) (b) (c) (d)

Figure 14: Adjacent a and c tiles results in a UET2 tiling

Having concluded that a a cannot be in the neighborhood of c, next consider the specific side lengths for c when $a + b = d$ that arise from the configuration in Figure 13b based on the knowledge that the vortex conditions require that a tile or tiles fit exactly between the dashed lines that extend toward the left of the union of the left edges of b and c in this picture. Note that one tile is too small to fill this space completely, four or more tiles are too large to fill this space completely, and in the case where three tiles exactly fill this space, a non-vortex tile must be in the middle of the group, forcing one of the three tiles to be a d as it is our only non-vortex. Using these facts, the following table enumerates all possibilities where they exist.

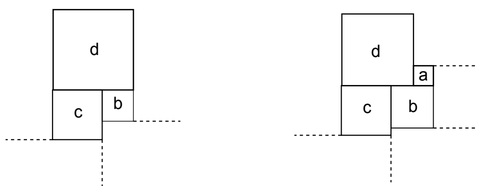
Therefore, when c has a tile b in its corona, the side lengths that must be considered are $c = 2a$ and $c = 3a$.

The final case to be considered is that in Figure 13c. Now, because no specific relationships for the side length of c can be established when it is surrounded by only d's, consider arrangement in Figure 13c where c's corona contains a tile other than only d. Without loss of generality, suppose that this non-d tile can be found along the right edge of c in Figure 13c. It is clear that this tile cannot be a c tile because the resulting tiling would not be

All possible combinations of two or three tiles	Can this combination fill the length in question exactly?	If yes, what does this imply about c?
a and b	no	
a and c	no	
a and d	yes	$b + c = a + d = 2a + b$ $\rightarrow c = 2a$
b and c	yes	$b + c = b + d$ $\rightarrow$ no new info
b and d	no	
c and d	no	
a and d and a	yes	$b + c = 2a + d = 3a + b$ $\rightarrow c = 3a$
a and d and b	no	
a and d and c	no	
b and d and b	no	
b and d and c	no	
c and d and c	no	

Table 4

unilateral. Lemma 2.4 implies that this tile cannot be a a tile. Then the non-d tile that must be found in c's corona is a b. Knowing that b must be a vortex, there are two arrangements that can result from this, shown in Figure 15:



(a) (b)

Figure 15

The arrangement in Figure 15a produces a contradiction because it implies $d = b + c > a + b = d$, so the arrangement in Figure 15b is the only viable UET4 possibility. This arrangement implies $b + c = a + d = 2a + b$, so $c = 2a$. Note that this c side length was already found in Table 4. Therefore, the only c side lengths that must be considered when $a + b = d$ are

1. $c = 2a$
2. $c = 3a$

along with the case in which c's corona contains only d tiles, in which case the only restriction is given by $b < c < a + b$.

Below is a summary of the side lengths considered when a and b are adjacent.

$c = a + b$		$d = a + b$
d is not a vortex	**d is a vortex**	
1. $d = a + b + c = 2a + 2b$ 2. $d = b + c = a + 2b$ 3. $d = a + c = 2a + b$ 4. $d = 2b + c = 3b + a$ 5. $d = 2a + c = 3a + b$	1. $d = 2b$ 2. $d = 3a$ 3. $d = 3b$	1. $c = 2a$ 2. $c = 3a$

3 Corona Construction

With the general UET4 problem having been effectively divided into subcases within which the problem can be appropriately examined, exhaustive lists of all possible coronas for the four sizes of tiles when a and b are adjacent can now be created. The process of constructing all possible coronas for a given case is begun by creating squares of the specified dimensions. Before beginning construction, it should be noted that, for any given tile, there exists at least one edge that extends into the skeleton of the tiling in no more than one direction, meaning it is compatible with the following figure:

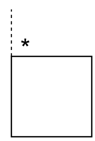


Figure 16

As neighborhoods are constructed, it is assumed that this necessary edge is the top edge of the tile in question.

The illustration of a partial example of corona construction is now presented so as to familiarize the reader with the general process used by examining a specific subcase. In order to illustrate the process used to create all a, b, c, and d coronas for a given set of side length proportions, a partial example is now outlined. Consider the case where a and b are adjacent, $c = a + b$, and $d = a + 3b$. The process is illustrated here by constructing all possible d coronas for this case, as these are the most complicated coronas to construct; it should be noted that c, b, and a coronas would also need to be constructed for this case. It is known that the arrangement in Figure 16 must appear in any corona, so the tiles that could be placed in the marked corner in that figure are first considered. An a, b, or c tile could be placed there, creating three branches shown in Figure 17 that will each be considered in turn.

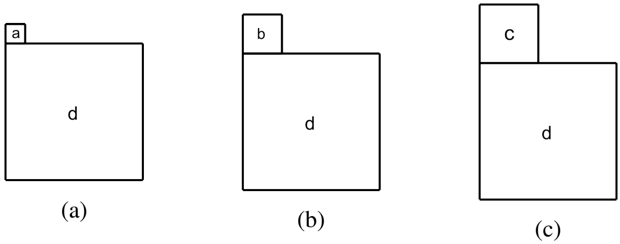


Figure 17: Three branches to consider for all possible d coronas

Take first the arrangement in Figure 17a; the corona is constructed by placing tiles around this center d in a clockwise direction. As the remaining length along the top edge of d is $3b$, placing an a or a b next would violate vortex conditions. Hence a c or a d can come next in the corona, creating two new branches shown below.

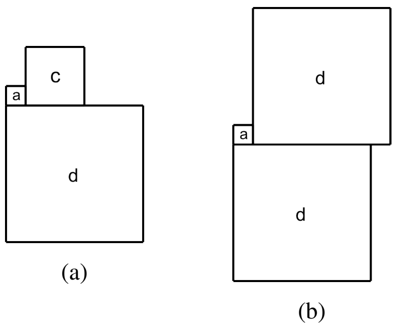


Figure 18: Two branches from Figure 17a

Consider the arrangement in Figure 18a. The length remaining along the top edge of d is $2b - a$, so neither an a nor a b can come next due to vortex conditions. A c is also not allowed by unilaterality. A tile d must come next, shown in Figure 19a. Now moving to the right edge of the center d tile, it is evident that a tile d cannot come next by unilaterality; again, vortex conditions say that a nor b can be in this next space either. A tile c must come next, shown in Figure 19b. There is now a distance of $2b$ along the remaining right edge of the center d tile, so it is again concluded that only a tile d could come next, seen in Figure 19c. Similar logic is employed to conclude that, continuing to move clockwise around the center tile, the remaining sides are covered by a c, then a d, then a c, then a d. This arrangement is shown in Figure 19d. However, this contradicts the vortex condition on the a tile, so this arrangement is invalid. Having exhausted all possibilities that could arise from the arrangement in Figure 18a, it is concluded that no viable coronas come from this branch, and attention is next given to the arrangement in Figure 18b. In order to minimize the creation of coronas that are identical up to cyclic permutations of their signature, the arrangement in Figure 18a is henceforth never considered along any edge of the center d tile.

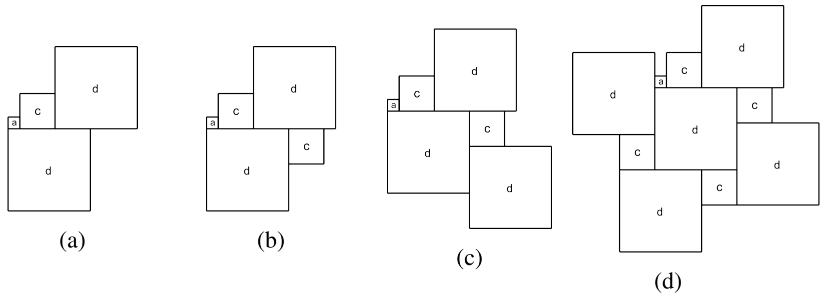


Figure 19: Successively filling the partial corona in Figure 19a

Corona construction proceeding from the branch in Figure 18b continues along the left edge of the lower d tile, moving clockwise as always. Given this arrangement, it is possible to satisfy vortex conditions if an a or b were placed next in the corona (note that the side length $b = 2a$ is required if this next tile is a b; an a is required along the remaining edge of this b in order to make it a vortex). A tile c could come next as well, but a tile d cannot by unilaterality. This leads to three additional branches, shown in Figure 20.

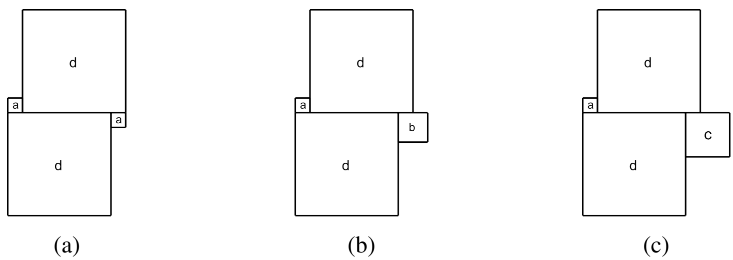


Figure 20: Three branches from Figure 18b

Consider now the arrangement in Figure 20a. Next, vortex conditions eliminate an a or b, so there are two options for the next tile in the corona, namely a c or a d as shown in Figure 21. Recall that the arrangement in Figure 21a contains on its right side a partial corona for which all possibilities were previously exhausted; this branch does not need to be reconsidered.

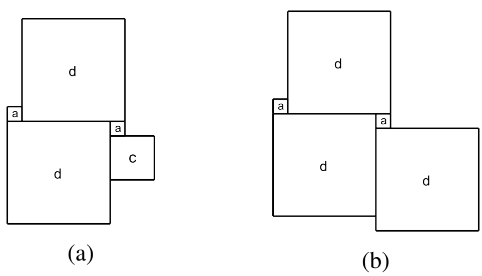


Figure 21: Two branches from Figure 20a

Consider the arrangement in Figure 21b. A d tile is the only one that could not come next in the sequence, by unilaterality. The three branches in Figure 22 result (note that the condition $b = 2a$ is again invoked in Figure 22b).

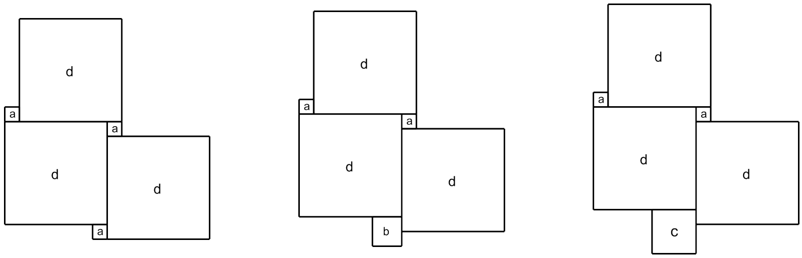


Figure 22: Three branches from Figure 21b

Consider the arrangement in Figure 22a. In choosing the next tile, vortex conditions eliminate a or b, so c or d could come next in the corona, but as placing a c in this position would lead to a cyclic permutation of an arrangement previously exhausted, only the case where a d comes next is considered. Following this d (seen in Figure 23a), a tile a, b, or c could be placed (d being disallowed by unilaterality). These three branches are shown in Figures 23b, 23c, and 23d.

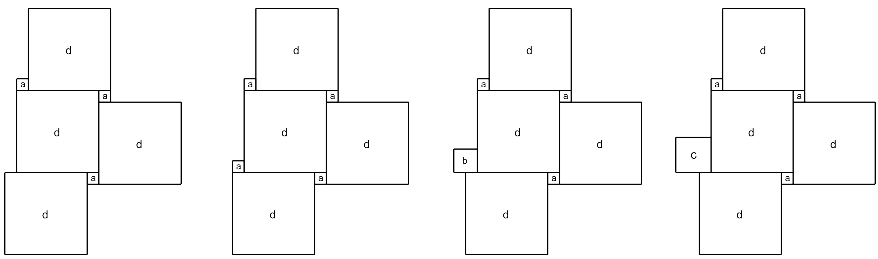


Figure 23: Partial corona $a.d.a.d.a.d$ and three resulting branches

As placing a c next in the partial corona of Figure 23b has been shown to be an impossible arrangement (see the branch from Figure 18a), only a d could complete this corona appropriately. This gives the complete corona signature $a.d.a.d.a.d.a.d$. Next, the vortex condition on the first a tile in Figure 23c would require that a combination of tiles fit exactly along the remaining distance $2a + 2b$ along the left edge. Tile lengths require that this include at least one d tile, which would violate the vortex conditions. Hence the arrangement in Figure 23c does not lead to a viable corona. Finally, the partial corona of Figure 23d can only be completed with a d, again violating the vortex condition on a; no viable coronas result from this arrangement. Now, all arrangements branching from that in Figure 23a have been exhausted. Figure 22b is the branch that should be returned to next.

The partial example outlined in this section is illustrated by a tree diagram in Figure 24; this shows branches of all possibilities considered along with which branches produce viable UET4 coronas and which are unfruitful. The tree diagram shows all branches that arise from placing an a tile in the asterisked position in Figure 16; those arising from placing a b or c in that position are not included in the diagram for brevity's sake.

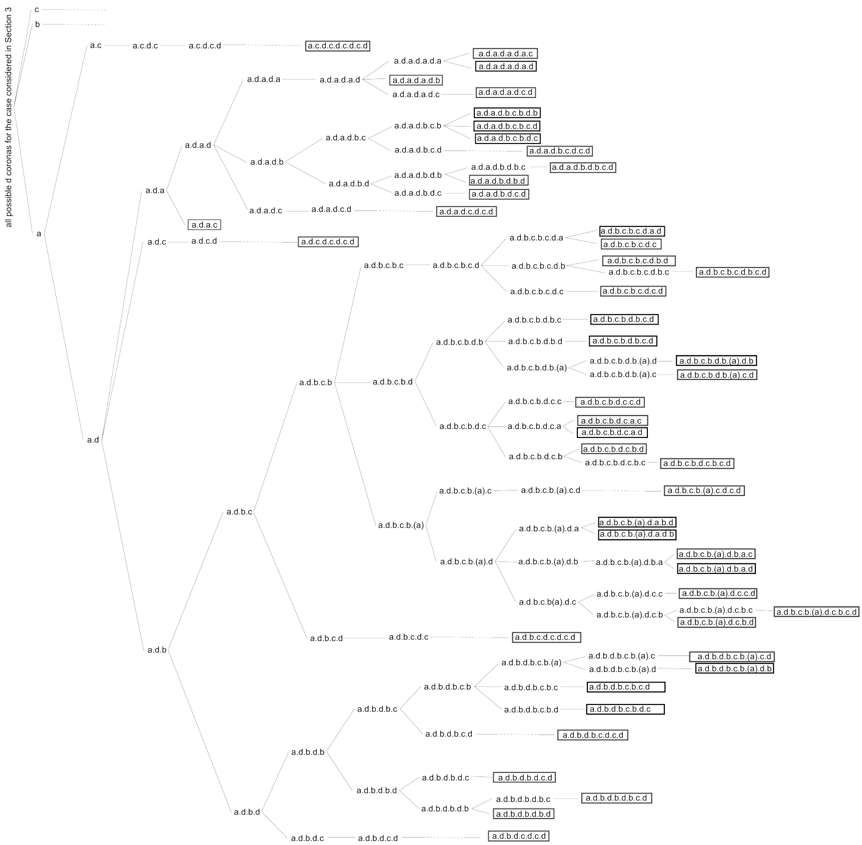


Figure 24: A tree diagram to accompany the illustration of corona construction found in Section 4. All d corona possibilities illustrated in Section 4 are pictured here; those omitted are not featured here either. Coronas in black rectangles are viable UET4 coronas; those in red rectangles (full or partial) do not lead to viable UET4 coronas. The dotted lines mark where a step or series of steps (for each of which only one possible tile could appear next) in the construction process have been omitted for brevity's sake.

4 Construction of the UET4 Tilings from Viable Coronas

Having compiled a list of all possible a, b, c, and d coronas for each of the cases when a and b are adjacent, it remains to determine which combinations of these coronas can be combined to generate a UET4 tiling. The extension of Schattschneider's method of finding coronas that correspond to a tiling is illustrated with an example. Consider the case where a and b are adjacent, $c = a + b$, and $d = 3a$. Using the process explained earlier, the corresponding set of viable coronas are:

a **coronas**	b **coronas**	c **coronas**	d **coronas**
$b.b.c.b$	$a.b.a.b.a.b.a.b$	$a.b.c.b.a.b.c.b$	$b.d.b.d.b.d.b.d$
$c.b.c.b$	$a.b.a.b.a.c.c$	$a.b.c.b.a.c.a.c$	$c.d.c.d.c.d.c.d$
$b.b.c.c$	$a.b.a.b.a.d.c$	$a.b.c.b.c.b.c.b$	$a.d.a.d.a.d.a.d$
$c.b.c.d$	$a.b.a.c.c.c$	$a.c.a.c.a.c.a.c$	$a.d.a.d.a.d.a.d$
$c.b.c.c$	$a.b.a.c.d.c$	$b.c.b.c.b.c.b.c$	$a.c.b.c.a.c.b.c$
$d.b.c.d$	$a.b.a.d.c.c$	$d.d.d.d$	$a.c.b.c.a.c.c.b$
$d.c.d.d$	$a.b.a.d.d.c$	$a.b.d.c.d.a.c$	$a.c.b.c.a.d.c.b$
$d.d.c.c$	$a.c.c.a.c.c$	$a.b.d.d.c.a.c$	$a.c.b.c.c.b.a.d$
$c.d.c.d$	$a.c.c.a.d.c$	$a.b.d.d.d.c$	$a.c.b.c.c.b.a.d$
$c.d.c.c$	$a.c.c.c.c$	$a.b.d.c.a.b.d$	$a.c.b.c.c.d.c.b$
$c.c.c.c$	$a.c.c.d.c$	$a.b.d.d.a.b.d$	$a.c.b.c.c.d.c.b$
$d.d.d.d$	$a.c.c.d.c$	$a.b.d.d.c.d$	$a.c.b.c.c.d.c.b$
	$a.c.d.c.a.b$	$a.b.c.b.c.b.a.c$	$a.c.b.d.b.c.a.d$
	$a.c.d.c.c$	$a.b.c.b.d.c.b$	$a.c.b.d.b.c.c.b$
	$a.c.d.d.c$	$a.b.c.b.d.d.b$	$a.c.b.d.b.d.c.b$
	$a.d.c.a.b.a.b$	$a.b.c.b.d.d.c$	$a.c.b.d.b.d.b.c$
	$a.d.c.a.c.c$	$a.b.c.b.d.c.d$	$a.c.b.d.c.b.a.d$
	$a.d.c.a.d.c$	$a.b.c.b.d.a.b.d$	$a.c.b.d.c.d.c.b$
	$a.d.c.c.a.b$	$a.b.c.b.d.a.d.b$	$a.c.c.b.a.c.b.c$
	$a.d.c.c.c$	$a.b.c.d.b.a.d.b$	$a.c.c.b.a.c.c.b$
	$a.d.c.d.c$	$a.b.c.d.b.a.b.d$	$a.c.c.b.a.d.a.d$
	$a.d.d.c.a.b$	$a.b.d.a.b.d.a.c$	$a.c.c.b.a.d.c.b$
	$a.d.d.c.c$	$a.b.d.a.c.a.b.d$	$a.c.c.d.c.b.a.d$
	$a.d.d.d.c$	$a.b.d.a.c.a.d.b$	$a.c.c.d.c.d.c.b$
	$c.c.c.c$	$a.b.d.a.d.b.a.c$	$a.d.a.d.a.d.c.b$
	$c.c.c.d$	$a.b.d.c.b.a.b.d$	$a.d.a.d.c.d.c.b$
	$c.d.c.d$	$a.b.d.c.b.a.d.b$	$a.d.c.b.a.d.c.b$
	$c.c.d.d$	$b.c.b.c.b.c.d$	$a.d.c.d.c.b.a.d$
	$c.d.d.d$	$b.c.b.c.b.d.d$	$a.d.c.d.c.d.c.b$
	$d.d.d.d$	$b.c.b.c.d.d$	
		$b.c.b.d.d.d$	
		$b.c.d.b.c.d$	
		$b.c.d.b.d.d$	
		$b.c.d.d.d$	
		$b.d.d.b.d.d$	
		$b.d.d.d.d$	
		$a.b.d.b.a.b.d$	
		$a.b.d.b.a.d.b$	

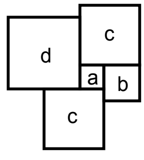


Figure 25: An a corona with signature $c.b.c.d$.

Choose the a corona $c.b.c.d$, as illustrated in Figure 25. Observe that because of the tiles neighboring the b tile in Figure 25, and because $\mathscr{T}$ is equitransitive, the b corona for

$\mathscr{T}$ must contain the partial corona $c.a.c$. Likewise, the c corona for $\mathscr{T}$ must contain the partial coronas $b.a.d$ and $d.a.b$ and the d corona for $\mathscr{T}$ must contain the partial corona $c.a.c$. Finding the partial b coronas that are compatible with our choice of a corona is easily automated. For example, to find the first partial b corona, find the first instance of b in the a corona, $c.b.c.d$. A partial corona signature for b would then be the letter cyclically preceeding this instance of b, followed by a (since this partial signature is taken from an a corona signature), followed by the letter cyclically following this instance of b, yielding the partial b corona signature $c.a.c$. This process can be repeated for each occurrence of b and for c and d as well. Note that partial corona signatures that are cyclic permutations of each other are considered equivalent, and so are reverse orderings. Performing such a search for partial b, c, and d corona signatures corresponding to the initial choice of a corona signature, $c.b.c.d$, yields the following.

A	**B**	**C**	**D**
a **corona**	**Partial** b **coronas**	**Partial** c **coronas**	**Partial** d **coronas**
$c.b.c.d$	$c.a.c$	$b.a.d$	$c.a.c$

Search the list of full b coronas for any which contain the partial corona $c.a.c$. In this example, after the search is performed, the five matching b coronas include $a.c.c.a.c.c$, $a.c.c.a.d.c$, $a.c.c.c.c$, $a.c.c.d.c$ and $a.c.d.d.c$. Use these coronas to create five corresponding 2-tuples of compatible a and b corona signatures (e.g. ($c.b.c.d$, $a.c.c.a.c.c$), ($c.b.c.d$, $a.c.c.a.d.c$), etc). For each new 2-tuple (x, y), add the b corona's corresponding partial a, c and d coronas to their respective columns as in Table 5.

	A	**B**	**C**	**D**	**E**
	(x, y) **Tuple**	**partial** a **coronas in** (x, y)	**partial** b **coronas in** (x, y)	**partial** c **coronas in** (x, y)	**partial** d **coronas in** (x, y)
1	($c.b.c.d$, $a.c.c.a.c.c$)	$c.b.c$	$c.a.c$	$a.b.c$ $b.a.d$	$c.a.c$
2	($c.b.c.d$, $a.c.c.a.d.c$)	$c.b.c$ $c.b.d$	$c.a.c$	$d.b.a$ $a.b.c$ $b.a.d$	$a.b.c$ $c.a.c$
3	($c.b.c.d$, $a.c.c.c.c$)	$c.b.c$	$c.a.c$	$a.b.c$ $c.b.c$ $b.a.d$	$c.a.c$
4	($c.b.c.d$, $a.c.c.d.c$)	$c.b.c$	$c.a.c$	$a.b.c$ $c.b.d$ $d.b.a$ $b.a.d$	$c.b.c$ $c.a.c$
5	($c.b.c.d$, $a.c.d.d.c$)	$c.b.c$	$c.a.c$	$a.b.d$ $b.a.d$	$c.b.d$ $c.a.c$

Table 5

While each of the b coronas in (x, y) contain the necessary partial b corona signature $c.a.c$, some may contain other partial a signatures that are not compatible with the original choice of a signature in (x, y). For example, the b corona of (x, y) in line 2, $a.c.c.a.d.c$, has the partial a corona $c.b.d$, which is not contained in the original a corona, $c.b.c.d$. So Line 2 is eliminated from further consideration.

Next, for each surviving tuple (x, y) in Table 5, search the list of full c coronas for those that contain the tuple's corresponding partial c coronas. For example, the tuple in line 4 has partial c coronas $a.b.d$, $d.b.c$, $c.b.a$ and $b.a.d$. All of these partial coronas are contained only in full c corona $a.b.c.b.d.a.b.d$. For each full c corona that is compatible with our 2-tuple (x, y), create a new 3-tuple (x, y, z) by appending compatible c corona signature, as in column **A** of Table 6. Also, list the partial a, b, c, and d coronas that are contained in (x, y, z) (columns **B - E**).

There are now nineteen viable (x, y, z) tuples. For each of these 3-tuples, check if all of the partial a and b coronas are contained in the full a and b coronas in the tuple. If not, remove that tuple. For example, consider the 3-tuple $(c.b.c.d, a.c.c.a.c.c, a.b.c.b.d.c.d)$ of Line 4. $a.c.d$ is listed as a partial b corona, but $a.c.d$ is not contained in this tuple's full b corona, which is $a.c.c.a.c.c$. Therefore, delete line 4. After performing this check for all of the tuples, there are only three tuples which pass the test, shown in Table 7.

Finally, search the list of full d coronas that contain all of the partial d coronas for each (x, y, z) 3-tuple to create a new list of 4-tuples (x, y, z, w) where w is a d corona that is compatible with the 3-tuple (x, y, z). For this example, this is Column A of Table 8. Add to this table columns containing the partial a, b, c, and d coronas contained in (x, y, z, w) which will be used to check the viability of (x, y, z, w) as before.

For each tuple, check if all of the partial a, b and c coronas are contained in the full a, b and c coronas in the tuple. For the tuple in line 1, this is not the case. Partial a corona $b.d.c$ is not contained in full a corona $c.b.c.d$. The tuple in line 2 passes the test.

At this point, from our original conditions and choice of a corona signature, there remains only one combination of a, b, c, and d coronas that may result in a tiling or tilings. When this process is automated and performed for all possible cases (e.g. $2a + b = d$)) and choices of a corona signatures, the following list of 4-tuples (x, y, z, w) is generated.

The process of constructing a tiling is demonstrated using a 4-tuple from Table 9. Consider the 4-tuple ($d.b.c.d$, $a.d.d.c.c$, $a.b.c.b.d.c.d$, $a.c.c.b.d.b.a.d$), for which $b = 2a$, $c = a + b$, and $d = 2a + b$. It is known that $b = 2a$ because it is a necessary condition for a c tile to have the corona $a.b.c.b.d.c.d$. The tiles and their coronas are displayed in Figure 26.

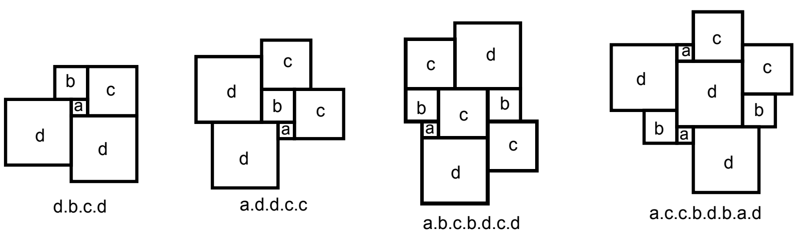


Figure 26: ($d.b.c.d$, $a.d.d.c.c$, $a.b.c.b.d.c.d$, $a.c.c.b.d.b.a.d$), for which $b = 2a$, $c = a + b$ and $d = 2a + b$

There are many ways to construct a tiling. The end goal is to create a patch that will tile the plane. For example, start with the d tile and its corona. Complete the coronas of the tiles which surround the d using reflections and rotations of the coronas in Figure 26. This

	A	**B**	**C**	**D**	**E**
	(x, y, z) **Tuple**	**partial** a **coronas in** (x, y, z)	**partial** b **coronas in** (x, y, z)	**partial** c **coronas in** (x, y, z)	**partial** d **coronas in** (x, y, z)
1	$(c.b.c.d,\ a.c.c.a.c.c,$ $a.b.c.d.b.a.b.d)$	$b.c.d$ $b.c.b$ $c.b.c$	$c.c.a$ $a.c.d$ $c.a.c$	$a.b.c$ $b.a.d$	$a.c.b$ $c.c.b$ $c.a.c$
2	$(c.b.c.d,\ a.c.c.a.c.c,$ $a.b.d.c.b.a.d.b)$	$b.c.b$ $b.c.d$ $c.b.c$	$a.c.c$ $a.c.d$ $c.a.c$	$a.b.c$ $b.a.d$	$c.c.b$ $a.c.b$ $c.a.c$
3	$(c.b.c.d,\ a.c.c.a.c.c,$ $a.b.c.d.b.a.d.b)$	bcb $b.c.d$ $c.b.c$	$a.c.d$ $a.c.c$ $c.a.c$	$a.b.c$ $b.a.d$	$b.c.c$ $a.c.b$ $c.a.c$
4	$(c.b.c.d,\ a.c.c.a.c.c,$ $a.b.c.b.d.c.d)$	$d.c.b$ $c.b.c$	$c.c.a$ $a.c.d$ $c.a.c$	$a.b.c$ $b.a.d$	$a.c.b$ $b.c.c$ $c.a.c$
5	$(c.b.c.d,\ a.c.c.a.c.c,$ $a.b.d.c.b.a.b.d)$	$d.c.b$ $b.c.b$ $c.b.c$	$c.c.a$ $a.c.d$ $c.a.c$	$a.b.c$ $b.a.d$	$b.c.a$ $b.c.c$ $c.a.c$
6	$(c.b.c.d,\ a.c.c.a.c.c,$ $a.b.c.b.d.a.b.d)$	$d.c.b$ $c.b.c$	$a.c.d$ $a.c.c$ $c.c.d$ $c.a.c$	$a.b.c$ $b.a.d$	$b.c.a$ $c.a.c$
7	$(c.b.c.d,\ a.c.c.d.c,$ $a.b.c.b.d.a.b.d)$	$d.c.b$ $c.b.c$	$a.c.d$ $a.c.c$ $c.c.d$ $c.a.c$	$a.b.c$ $c.b.d$ $d.b.a$ $b.a.d$	$b.c.a$ $c.b.c$ $c.a.c$
8	$(c.b.c.d,\ a.c.d.d.c,$ $a.b.c.d.b.a.d.b)$	$b.c.b$ $b.c.d$ $c.b.c$	$a.b.d$ $b.a.d$	$a.c.d$ $a.c.c$	$b.c.c$ $a.c.b$ $c.b.d$ $c.a.c$
9	$(c.b.c.d,\ a.c.d.d.c,$ $a.b.d.d.c.d)$	$d.c.b$ $c.b.c$	$a.b.d$ $b.a.d$	$a.c.d$	$a.c.c$ $d.c.c$ $b.c.d$ $c.b.d$ $c.a.c$
10	$(c.b.c.d,\ a.c.d.d.c,$ $a.b.d.b.a.b.d)$	$b.c.d$ $b.c.b$ $c.b.c$	$a.b.d$ $b.a.d$	$a.c.d$	$a.c.b$ $b.c.b$ $c.b.d$ $c.a.c$
11	$(c.b.c.d,\ a.c.d.d.c,$ $a.b.d.c.b.a.b.d)$	$b.c.d$ $b.c.b$ $c.b.c$	$a.b.d$ $b.a.d$	$a.c.c$ $a.c.d$	$a.c.b$ $b.c.c$ $c.b.d$ $c.a.c$
12	$(c.b.c.d,\ a.c.d.d.c,$ $a.b.d.d.a.b.d)$	$b.c.d$ $c.b.c$	$a.b.d$ $b.a.d$	$a.c.d$	$a.c.d$ $b.c.d$ $c.b.d$ $c.a.c$
13	$(c.b.c.d,\ a.c.d.d.c,$ $a.b.d.c.a.b.d)$	$b.c.d$ $c.c.b$ $c.b.c$	$a.b.d$ $b.a.d$	$a.c.d$	$a.c.b$ $b.c.c$ $c.b.d$ $c.a.c$
14	$(c.b.c.d,\ a.c.d.d.c,$ $a.b.d.c.b.a.d.b)$	$b.c.b$ $b.c.d$ $c.b.c$	$a.b.d$ $b.a.d$	$a.c.c$ $a.c.d$	$c.c.b$ $a.c.b$ $c.b.d$ $c.a.c$
15	$(c.b.c.d,\ a.c.d.d.c,$ $a.b.d.b.a.d.b)$	$b.c.b$ $b.c.d$ $c.b.c$	$a.b.d$ $b.a.d$	$a.c.d$	$b.c.b$ $a.c.b$ $c.b.d$ $c.a.c$
16	$(c.b.c.d,\ a.c.d.d.c,$ $a.b.d.a.c.a.b.d)$	$b.c.d$ $d.c.c$ $c.c.b$ $c.b.c$	$a.b.d$ $b.a.d$	$a.c.d$	$b.c.a$ $c.b.d$ $c.a.c$
17	$(c.b.c.d,\ a.c.d.d.c,$ $a.b.c.b.d.a.b.d)$	$d.c.b$ $c.b.c$	$a.b.d$ $b.a.d$	$d.c.a$ $a.c.c$ $c.c.d$	$b.c.a$ $c.b.d$ $c.a.c$
18	$(c.b.c.d,\ a.c.d.d.c,$ $a.b.c.d.b.a.b.d)$	$b.c.d$ $b.c.b$ $c.b.c$	$a.b.d$ $b.a.d$	$a.c.d$ $a.c.c$	$a.c.b$ $c.c.b$ $c.b.d$ $c.a.c$
19	$(c.b.c.d,\ a.c.d.d.c,$ $a.b.d.a.b.d.a.c)$	$b.c.d$ $d.c.c$ $c.c.b$ $c.b.c$	$a.b.d$ $b.a.d$	$a.c.d$	$b.c.a$ $c.b.d$ $c.a.c$

Table 6

	A	**B**	**C**	**D**	**E**
	(x, y, z) **Tuple**	**partial** a **coronas in** (x, y, z)	**partial** b **coronas in** (x, y, z)	**partial** c **coronas in** (x, y, z)	**partial** d **coronas in** (x, y, z)
1	($c.b.c.d$, $a.c.c.d.c$, $a.b.c.b.d.a.b.d$)	$d.c.b$ $c.b.c$	$a.c.d$ $a.c.c$ $c.c.d$ $c.a.c$	$a.b.c$ $c.b.d$ $d.b.a$ $b.a.d$	$b.c.a$ $c.b.c$ $c.a.c$
2	($c.b.c.d$, $a.c.d.d.c$, $a.b.d.d.c.d$)	$d.c.b$ $c.b.c$	$a.b.d$ $b.a.d$	$a.c.d$	$a.c.c$ $d.c.c$ $b.c.d$ $c.b.d$ $c.a.c$
3	($c.b.c.d$, $a.c.d.d.c$, $a.b.d.d.a.b.d$)	$b.c.d$ $c.b.c$	$a.b.d$ $b.a.d$	$a.c.d$	$a.c.d$ $b.c.d$ $c.b.d$ $c.a.c$

Table 7

	A	**B**	**C**	**D**	**E**
	(a, b, c, d) **Tuple**	**Tuple's partial** a **coronas**	**Tuple's partial** b **coronas**	**Tuple's partial** c **coronas**	**Tuple's partial** d **coronas**
1	($c.b.c.d$, $a.c.c.d.c$, $a.b.c.b.d.a.b.d$, $a.c.b.c.a.c.c.b$)	$d.c.b$ $c.b.c$ $b.d.c$ $c.d.c$	$a.c.d$ $a.c.c$ $c.c.d$ $c.a.c$ $c.d.a$ $c.d.c$	$a.b.c$ $c.b.d$ $d.b.a$ $b.a.d$ $a.d.b$ $a.d.c$ $c.d.b$	$b.c.a$ $c.b.c$ $c.a.c$
2	($c.b.c.d$, $a.c.c.d.c$, $a.b.c.b.d.a.b.d$, $a.c.b.c.a.c.b.c$)	$d.c.b$ $c.b.c$ $c.d.c$	$a.c.d$ $a.c.c$ $c.c.d$ $c.a.c$ $c.d.c$	$a.b.c$ $c.b.d$ $d.b.a$ $b.a.d$ $a.d.b$	$b.c.a$ $c.b.c$ $c.a.c$

Table 8

results in the patch illustrated in Figure 27.

4-tuple	**Proportions**
$(d.d.d.d$, $c.d.c.d$, $b.d.d.b.d.d$, $a.d.a.d.c.b.c.d)$*	$d = c + b$
$(d.d.d.d$, $d.c.c.c$, $b.c.b.c.b.d.d$, $a.d.a.d.b.c.b.c.d)$*	$d = c + b$
$(c.d.c.d$, $d.d.d.d$, $a.d.d.a.d.d$, $a.c.d.b.d.b.d.c)$*	$d = c + a$
$(d.d.d.d$, $c.d.c.d$, $d.d.d.b.d$, $a.d.c.d.c.b.c.d)$*	$d = c + b$
$(c.d.c.d$, $d.d.d.d$, $a.d.d.d.d$, $a.c.d.c.d.b.d.c)$*	$d = c + a$
$(c.c.c.d$, $d.d.d.d$, $a.c.a.c.a.d.d$, $a.c.d.b.d.b.d.c)$*	$d = c + a$
$(d.d.c.c$, $c.d.c.d$, $a.c.a.d.b.d.d$, $a.c.d.a.c.b.c.d)$*	$b = 2a$; $d = c + a$
$(d.b.c.d$, $a.d.d.d.c$, $a.b.d.a.b.d$, $a.c.b.d.b.d.b.a.d)$	$d = 2a + 2b$
$(d.b.c.d$, $a.d.d.d.c$, $a.b.d.a.b.d$, $a.c.b.d.b.d.a.b.d)$	$d = 2a + 2b$
$(d.b.c.d$, $a.d.c.a.d.c$, $a.b.d.d.d$, $a.c.d.c.d.c.b.a.d)$	$b = 3a$; $d = 2a + c$
$(c.b.c.d$, $a.c.d.c.c$, $a.b.c.b.d.a.b.d$, $a.c.b.c.a.c.b.c)$	$d = 3a$
$(d.b.c.d$, $a.d.c.d.c$, $a.b.d.d.b.d$, $a.c.b.a.d.c.b.c.d)$	$d = a + 2b$
$(d.b.c.d$, $a.d.d.d.c$, $a.b.d.a.b.d$, $a.c.b.d.b.d.b.a.d)$	$d = a + 2b$
$(d.b.c.d$, $a.d.d.d.c$, $a.b.d.a.b.d$, $a.c.b.d.b.d.a.b.d)$	$d = a + 2b$
$(d.b.c.d$, $a.d.c.c.c$, $a.b.d.a.b.d$, $a.c.b.a.d.a.c.b.a.d)$	$d = a + 2b$
$(c.b.c.b$, $a.c.d.c.c$, $a.b.d.d.b.c.b$, $b.c.d.c.b.c.d.c)$	$d = a + 2b$
$(b.b.c.c$, $a.b.a.c.d.c$, $a.b.d.d.b.a.c$, $b.c.d.c.b.c.d.c)$	$d = a + 2b$
$(b.c.b.c$, $a.c.d.d.c$, $a.b.d.d.d.b$, $b.c.d.c.d.c.b.d)$	$d = a + 2b$
$(b.c.d.d$, $a.d.d.c.c$, $a.b.c.b.d.d$, $a.c.d.c.b.d.a.b.d)$	$b = 2a$; $d = a + 2b$
$(c.b.c.b$, $a.c.d.d.c$, $a.b.d.d.d.b$, $b.c.d.c.d.b.c.d)$	$d = a + 2b$
$(d.b.c.d$, $a.d.c.a.d.c$, $a.b.d.b.a.d$, $a.c.a.d.a.b.c.b.a.d)$	$d = a + 2b$
$(d.b.c.d$, $a.d.d.c.c$, $a.b.c.b.d.c.d$, $a.c.c.b.d.a.b.d)$	$d = 2a + b$
$(d.b.c.d$, $a.d.c.a.d.c$, $a.b.d.c.d.d$, $a.d.c.a.d.c.c.b)$	$b = 2a$; $d = 2a + b$
$(c.b.c.b$, $a.c.d.d.c$, $a.b.d.b.a.b.d.b$, $b.c.b.d.b.c.b.d)$	$d = 2a + b$
$(d.b.c.d$, $a.d.c.a.d.c$, $a.b.d.c.d.d$, $a.c.d.c.c.b.a.d)$	$b = 2a$; $d = 2a + b$
$(d.b.c.d$, $a.d.c.a.d.c$, $a.b.d.d.d$, $a.c.d.c.d.c.b.a.d)$	$d = 2a + b$
$(d.b.c.d$, $a.d.d.c.c$, $a.b.c.d.b.c.d$, $a.c.c.b.d.a.b.d)$	$b = 2a$; $d = 2a + b$
$(d.b.c.d$, $a.d.d.d.c$, $a.b.d.a.b.d$, $a.c.b.d.b.d.b.a.d)$	$d = 2a + b$
$(d.b.c.d$, $a.d.d.c.c$, $a.b.c.b.d.c.d$, $a.c.c.b.d.b.a.d)$	$b = 2a$; $d = 2a + b$
$(c.b.c.d$, $a.c.c.a.c.c$, $a.b.c.b.a.d.d$, $a.c.d.c.a.c.d.c)$	$d = 2a + b$
$(d.b.c.d$, $a.d.d.d.c$, $a.b.d.a.b.d$, $a.c.b.d.b.d.b.a.d)$	$d = 2a + b$
$(c.b.c.d$, $a.c.d.d.c$, $a.b.d.b.a.d.d$, $a.c.d.b.c.b.d.c)$	$d = 2a + b$
$(c.b.c.c$, $a.c.d.d.c$, $a.b.d.b.a.c.a.c$, $b.c.b.d.b.c.b.d)$	$d = 2a + b$
$(c.b.c.d$, $a.c.d.d.c$, $a.b.d.a.b.d.d$, $a.c.d.c.b.d.b.c)$	$d = 2a + b$
$(c.b.c.d$, $a.c.d.d.c$, $a.b.d.a.b.d.d$, $a.c.d.b.c.d.b.c)$	$d = 2a + b$
$(c.b.c.d$, $a.c.d.d.c$, $a.b.d.d.d$, $a.c.d.c.d.c.b.d.c)$**	$d = 2b + a$
$(c.b.c.d.$, $a.c.d.d.c.$, $a.b.d.c.d.d.$, $a.c.d.c.c.b.d.c)$**	$d = 2a + b$
$(d.b.c.d.$, $a.d.c.a.d.c.$, $a.b.d.b.a.d.d.$, $a.c.d.a.b.c.b.a.d)$**	$d = 2a + b$
$(c.d.c.d.$, $c.c.d.d.$, $a.d.b.c.d.d.$, $a.c.d.c.c.d.b.c)$**	$d = 2a + b$
$(c.d.c.d.$, $c.c.d.d.$, $a.d.b.c.b.d.d.$, $a.c.d.b.c.d.b.c)$**	$d = 2a + b$

Table 9: All 4-tuples generated when $a + b = c$. Note that all marked with (*) were already found in Section 2.1. All marked with (**) cannot be extended to create a tiling of the plane.

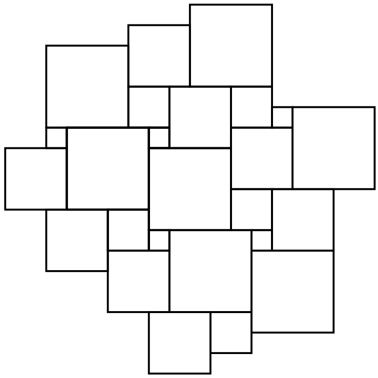


Figure 27: A second layer of the d corona for ($d.b.c.d$, $a.d.d.c.c$, $a.b.c.b.d.c.d$, $a.c.c.b.d.b.a.d$).

By adding tiles in a similar manner, such that the coronas are reflections and rotations of those in Figure 26, as well as deleting tiles where necessary, one will easily find a patch which can tile the plane unilaterally and equitransitively using translations. Such a patch is illustrated in Figure 28.

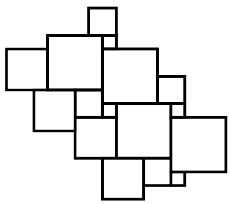


Figure 28: A patch which will tile the plane unilaterally and equitransitively by way of translations.

However, for some 4-tuples, it will soon become clear that no tiling is possible. For example, consider the 4-tuple ($c.b.c.d$, $a.c.d.d.c$, $a.b.d.d.d$, $a.c.d.c.d.c.b.d.c$), for which $c = a + b$ and $d = a + 2b$. If one attempts to expand on the d corona in a similar manner as above - by completing incomplete coronas while adhering to the ordering prescribed by the 4-tuple - one will encounter the patch in Figure 29.

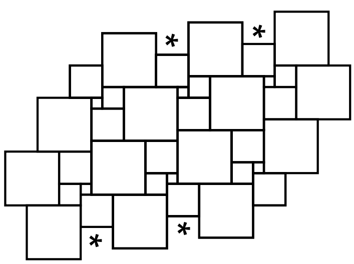


Figure 29: A patch which cannot be extended for 4-tuple ($c.b.c.d$, $a.c.d.d.c$, $a.b.d.d.d$, $a.c.d.c.d.c.b.d.c$), for which $c = a + b$ and $d = a + 2b$.

The (*) represent problem areas. To adhere to equitransitivity, one must place a d tile in these spots. Obviously, this is impossible. For each of the five tilings marked with (**) in Table 9, a patch which could not be extended was inevitable.

5 a and b are not adjacent

In this case, there are six possible a coronas. These are illustrated in Figure 30. By replacing the a tiles with b tiles in Figure 30, it is clear that there are also exactly six b coronas when a and b are not adjacent. Lemma 5.1 eliminates one of these six subcases of a coronas and one of these six subcases of b coronas from consideration in UET4 tilings.

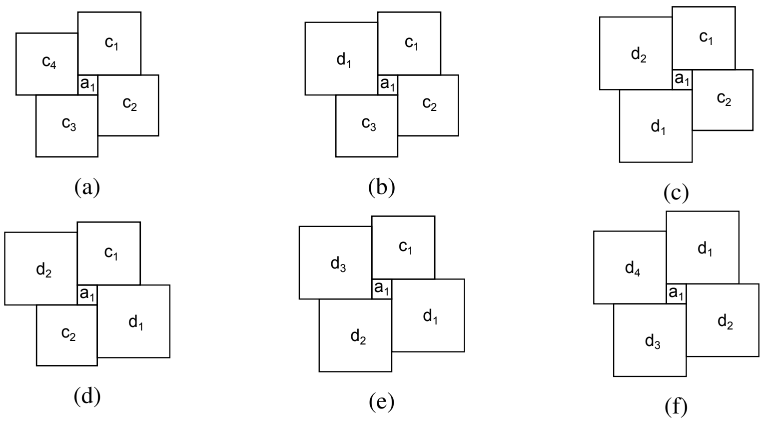


Figure 30: All possible a coronas when a and b are not adjacent

Lemma 5.1. *Let $\mathscr{T}$ be a UET4 tiling in which a and b are not adjacent. Then a coronas and b coronas cannot contain only c tiles.*

Proof. Suppose that the a corona can contain only c tiles as shown in Figure 31a below. At least one corner formed by two tiles c must contain a non-a tile; otherwise the resulting tiling would be UET2. Without loss of generality, suppose that this required corner is that marked by the asterisk in Figure 31a; it will be determined which tiles can be placed in the corner marked by the asterisk. Were a b tile to be placed here, then this b tile would overhang past the right edge of c_1. Since every b tile is a a vortex, the length of this overhang must be covered exactly by a tile (or tiles), and the only tile that can cover this length $b - a$ while maintaining the appropriate relative side lengths of a, b, c, and d is an a tile. However, this would contradict a and b not being adjacent. Therefore a d tile must fill this space as shown in Figure 31b. Next, it is determined which tiles could be placed in the corner marked by the asterisk in FIgure 31b. Were an a tile to be placed here, there would be a d tile in its corona and the tiling would cease to be equitransitive by the assumption that a's corona contained only c tiles. A c tile cannot fill the asterisked corner by unilaterality, so the two cases shown in Figures 31c and 31d must be considered, starting with Figure 31c.

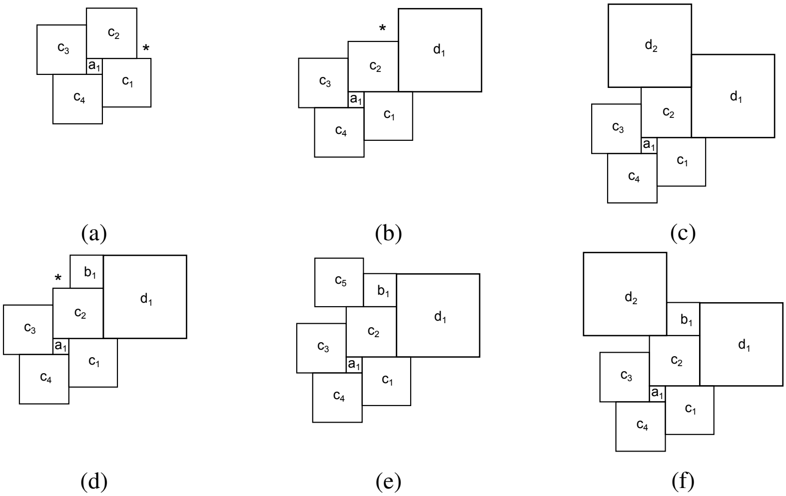


Figure 31: Progressively building all possible c coronas

The vertical distance remaining along the left edge of tile c_2 in this figure is of length a; hence the only tile that could appropriately fill this remaining edge length is an a tile. However, the corona of this new a tile would contain a d tile, contrary to hypothesis. Therefore the arrangement of tiles in Figure 31c does not give rise to a UET4 tiling.

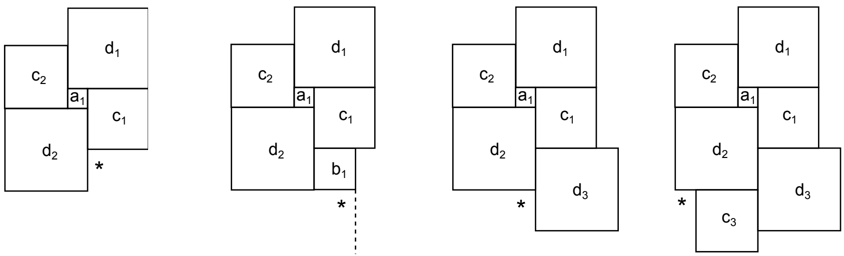
Next consider case illustrated in Figure 31d. Note that, as pictured, the top edge of tile b_1 must line up with the top edge of d_1. Were this not the case, either b_1 would cease to be a vortex or would be forced to have an a tile in its neighborhood. Thus, an a tile cannot be placed in the corner marked by the asterisk because a and b cannot be adjacent; neither can a b tile be placed there by the unilaterality condition. This leaves the two cases shown in Figure 31e and 31f. In both of these cases, the remaining vertical distance along the left edge of c_2 is of length a, so the only tile that could fill this space is an a tile. However, the distance that the bottom edge of c_5 in Figure 31e and the bottom edge of d_2 in Figure 31f hang over the left edge of c_2 is, in both cases, strictly greater than the length a. Hence an a tile placed along the remaining left edge of c_2 would not be a vortex. Therefore the arrangement of tiles found in Figure 31d does not give rise to any UET4 tilings.

A nearly identical argument shows that a b tile cannot be surrounded by only c tiles. □

Because neither an a tile nor a b tile can have a corona containing only c tiles, then each of their coronas must contain a d tile. An immediate corollary to this is that each d corona must contain at least one a tile and at least one b tile. Lemma 5.1 implies that there are five possible a coronas and five possible b coronas when a and b are not adjacent.

Lemma 5.1 illustrates the analysis of only one possible a corona, but there are 5 more a coronas to consider, 6 more b coronas, and several possible c and d coronas to consider. Because each possible corona involves exhaustive examination, it would be impractical to present such an analysis in a short article. However, the following example illustrates the methodology used to decide if a given corona is viable. Consider the a corona of Figure 30d. This corona is reprinted in Figure 32a. Proceed by constructing all possible d coronas that arise from this arrangement by placing tiles along the edges of the tile d_2 in Figure 32a moving in a clockwise direction. There are three subcases here to consider:

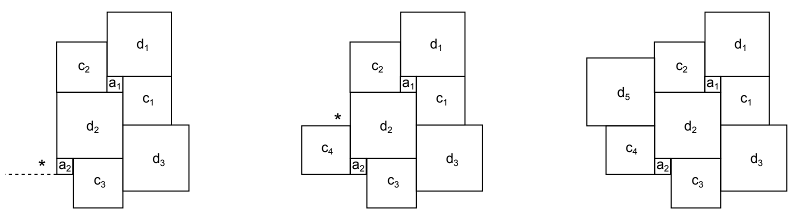
$d > a + c$, $d = a + c$, and $d < a + c$. In this example, only the subcase $d = a + c$ is demonstrated. To begin enumeration of all possible d coronas that can arise from the arrangement in Figure 32a, first determine which tiles could be placed in the corner marked with an asterisk. An a tile cannot be placed there because the vortex restriction on a tiles would imply that $c = d$. A c tile cannot be placed there by unilaterality. This gives us two options to consider: a tile b_1 can be placed there or a tile d_3 can be placed there. These two options are shown in Figures 32b and 32c. Note that for the case shown in Figure 32b, the vortex condition on b tiles and the fact that a tiles and b tiles cannot be adjacent requires that $a + d = b + c$.



(a) (b) (c) (d)

Figure 32

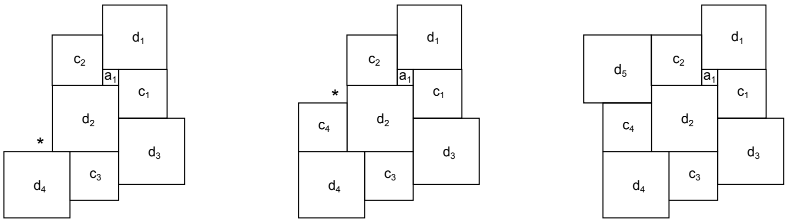
In a full analysis, both of the arrangements in 32b and 32c above need to be considered, but for the purposes of this example, consider only how to fill in the asterisked corner in Figure 32c. A d tile cannot be placed there by unilaterality; however, an a, a b, or a c could be placed there under the appropriate conditions. Each of these options needs to be considered. Examine the arrangement, shown in Figure 32d, where a tile c_3 fills the asterisked corner. Again, it must be determined which tiles can be placed in the asterisked corner in Figure 32d. Were a tile b placed there, the condition $d = a + c$ implies that the top edge of b will have overhang past the left edge of d_2; then in order for b's vortex condition to be satisfied, a tile a would have to be placed along the remaining top edge of b. This contradicts a and b not being adjacent. A c tile cannot be placed in the asterisked corner by unilaterality. This leaves two options to consider. The case an a tile in this position is seen in Figure 33a, and the case where a d tile in this position is seen in Figure 34a.



(a) (b) (c)

Figure 33

Consider first the arrangement shown in Figure 33a. By equitransitivity of a, a tile c_4 must be placed in the asterisked corner so that its corona will match that of a_1. This is shown in Figure 33b. The asterisked corner Figure 33b can only be filled by a tile d_4, shown in Figure 33c; a and b are not allowed there by the vortex conditions, and c is not allowed by unilaterality. Then Figure 33c shows a completed d corona. However, since this d corona contains no b tile, this will not result in a UET4 tiling as a result of a corollary to Lemma 5.1 and is hence not a viable d corona.



(a) (b) (c)

Figure 34

Next, considering the arrangement shown in Figure 34a, it must be determined which tiles can be placed in the corner marked by an asterisk. Vortex conditions prohibit an a or b from being placed there, and a d tile is also not allowed by unilaterality. Then the only option is to place a tile c_4 in this position, shown in Figure 34b. The asterisked corner in this figure can only be filled by a tile d_5, as shown in Figure 34c; a and b are not possible by vortex conditions and c is not possible by unilaterality. Figure 34c shows a complete d corona. However, because this corona does not contain a b tile, it is not compatible with a UET4 tiling.

The next step would be to consider the possibilities when an a tile or a b tile are placed in the position occupied by c_3 in Figure 32d. The method continues in this fashion, enumerating all possible tiles that can be placed in a location, moving around a d tile, creating a branching list of all d coronas and weeding out coronas that are known to be impossible under our constraints. This is done for the three subcases $c > a+c$, $d = a+c$, and $d < a+c$ for each of the five possible a coronas in Figures 30b-30f, and it seen that $d = a + c$ is the only case that yields viable d coronas. It should also be noted that in the case where an a tile is surrounded by three c tiles and one d tile, as shown in Figure 30b, the same method used to build around a c tile instead of a d tile. It should also be noted that, at times, it is necessary to specify certain side lengths for b and c tiles in terms of side lengths of smaller tiles in order for certain arrangements to be viable. This allows for further flexibility in corona construction and ensures that all possible potentially viable coronas are found.

Summarizing, the criteria used throughout this method are as follows:

1. Equitransitivity of the tiling $\mathscr{T}$.

   - When an a tile is placed in the corona of a larger tile, it is possible to continue building around the larger tile using the knowledge that every a neighborhood must be identical to that of the already established a_1 corona.
   - It is possible that at times the only option is to place a tile within an a tile's neighborhood that makes it incompatible with the original a_1 corona. In this case, the method can be ended on this branch, as it will not yield any viable c or d coronas.

2. Unilaterality of the tiling $\mathscr{T}$.
3. Vortex conditions on a and b tiles.
4. The requirement that a tiles and b tiles are not adjacent.
5. Relative sizes of tile side lengths: $a < b < c < d$.
6. Each d corona must contain at least one a tile and at least one b tile.

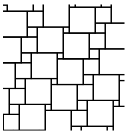
This method essentially begins with one of the five viable a coronas shown in Figure 30 and then employs the corona construction algorithm outlined previously in Section 3, building around a tile in the a corona (specifically, around a d tile in Figures 30c, 30d, and 30e or around a c tile in Figures 30a and 30b) until either a contradiction is reached for a particular branch or a full corona is reached. Once the construction process has been completed building around the chosen tile in the corona of the original a tile, one is left with an exhaustive list of all d coronas (or c coronas, depending on the a corona from which construction began) that are compatible with the original a corona. Then, using equitransitivity, the coronas of all new a and d tiles (or c tiles, again depending on the case) can be completed, expanding the patch until one can either establish that no UET4 tiling can result (due to failure of equitransitivity, overlapping or gaps between tiles, contradiction of vortex conditions, etc.) or until full b and c coronas (or b and d coronas) are found. Note that multiple tilings may result from the same set of a and d coronas (or a and c coronas), as there may be multiple ways to tile the original patch using these coronas. Table 10 lists the eleven UET4 tilings found using this method and the necessary side length proprtions required for the tiling to be generated.

(a **corona,** b **corona,** c **corona,** d **corona**)	**Side Relations**
$(d.d.d.d,\ c.d.c.d,\ b.d.d.b.d.d,\ a.d.a.d.c.b.c.d)$	$c = a + b;\ d = c + b$
$(d.d.d.d,\ d.c.c.c,\ b.c.b.c.b.d.d,\ a.d.a.d.c.b.c.d)$	$c = a + b;\ d = c + b$
$(c.d.c.d,\ d.d.d.d,\ a.d.d.a.d.d,\ a.c.d.b.d.b.d.c)$	$c = a + b;\ d = c + a$
$(d.d.c.c,\ c.d.c.d,\ b.d.d.a.c.a.d,\ a.d.a.c.d.c.b.c)$	$b = 2a;\ d = c + a$
$(d.d.d.d,\ c.d.c.d,\ d.d.d.b.d,\ a.d.c.d.c.b.c.d)$	$c = a + b;\ d = c + b$
$(c.d.c.d,\ d.d.d.d,\ a.d.d.d.d,\ a.c.d.c.d.b.d.c)$	$c = a + b;\ d = c + a$
$(c.d.c.d,\ c.d.c.d,\ a.d.b.d.d.d,\ b.c.a.c.d.c.d.c)$	$b = 2a;\ d = c + a$
$(c.c.c.d,\ d.d.d.d,\ a.c.a.c.a.d.d,\ a.c.d.b.d.b.d.c)$	$c = a + b;\ d = c + a$
$(d.d.c.c,\ c.d.c.d,\ a.c.a.d.b.d.d,\ a.c.d.a.c.b.c.d)$	$b = 2a;\ c = a + b;\ d = c + a$
$(c.d.c.d,\ c.c.d.d,\ a.d.b.c.b.d.d,\ a.c.b.d.c.b.d.c)$	$b = 2a;\ c = 2b;\ d = c + a$
$(c.d.c.d,\ c.c.d.d,\ b.c.b.d.d.a.d,\ a.c.d.c.b.d.b.c)$	$b = 2a;\ d = c + a$

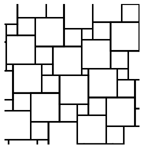
Table 10

Illustrations of the eleven tilings when a and b are not adjacent can be seen in the final section. This concludes the case where a tiles and b tiles are not allowed to be adjacent.

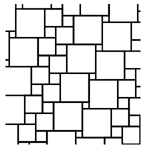
6 The 39 UET4 Tilings



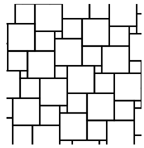
$(d.d.d.d, c.d.c.d, b.d.d.b.d.d, a.d.a.d.c.b.c.d)$
$c = a + b$
$d = c + b$



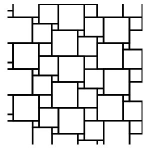
$(d.d.d.d, d.c.d.c, b.d.d.d.d, a.d.c.d.c.b.c.d)$
$c = a + b$
$d = c + b$



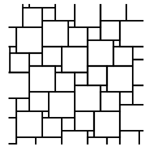
$(d.d.d.d, c.c.c.d, b.c.b.d.d.b.c, a.d.c.b.c.d.a.d)$
$c = a + b$
$d = b + c$



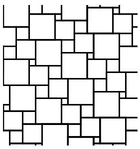
$(c.d.c.d, d.d.d.d, a.d.d.d.d, a.c.d.c.d.b.d.c)$
$c = a + b$
$d = c + a$



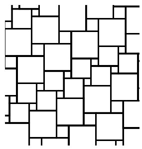
$(c.c.d.d, c.d.c.d, a.c.a.d.b.d.d, a.d.a.c.d.c.b.c)$
$b = 2a$
$d = c + a$



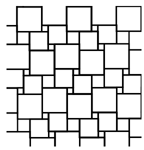
$(c.d.c.d, c.d.c.d, a.d.b.d.d.d, a.c.d.c.d.c.b.c)$
$b = 2a$
$d = c + a$



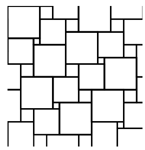
$(c.d.c.d, c.c.d.d, a.d.b.c.b.d.d, a.c.d.c.b.d.b.c)$
$b = 2a$
$d = c + a$



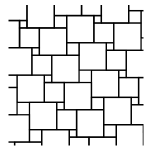
$(c.c.c.d, d.d.d.d, a.c.a.c.a.d.d, a.c.d.b.d.b.d.c)$
$c = a + b$
$d = c + a$



$(d.d.c.c, c.d.c.d, a.c.a.d.b.d.d, a.c.d.a.c.b.c.d)$
$b = 2a$
$c = a + b$
$d = c + a$

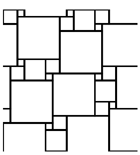


$(c.d.c.d, c.c.d.d, a.d.b.c.b.d.d, a.c.b.d.c.b.d.c)$
$b = 2a$
$c = 2b$
$d = c + a$

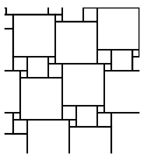


$(c.d.c.d, d.d.d.d, a.d.d.a.d.d, a.c.d.b.d.b.d.c)$
$c = a + b$
$d = c + a$

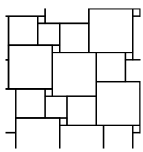
Below are the 28 tilings when a and b are adjacent. For all, $c = a + b$. $d = a + b$ does not generate any tilings.



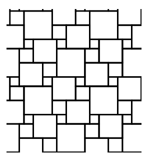
$(d.b.c.d, a.d.d.d.c, a.b.d.a.b.d, a.c.b.d.b.d.a.b.d)$
$d = 2a + 2b$



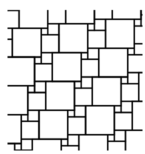
$(d.b.c.d, a.d.d.d.c, a.b.d.a.b.d, a.c.b.d.b.d.b.a.d)$
$d = 2a + 2b$



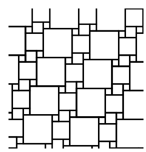
$(d.b.c.d, a.d.c.a.d.c, a.b.d.d.d, a.c.d.c.d.c.b.a.d)$
$b = 3a$
$d = 2a + c$



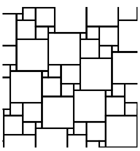
$(c.b.c.d, a.c.d.c.c, a.b.c.b.d.a.b.d, a.c.b.c.a.c.b.c)$
$d = 3a$



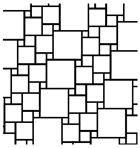
$(d.b.c.d, a.d.c.d.c, a.b.d.d.b.d, a.c.b.a.d.c.b.c.d)$
$d = a + 2b$



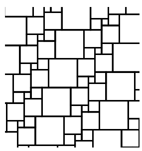
$(d.b.c.d, a.d.c.a.d.c, a.b.d.a.b.d, a.c.b.a.d.a.c.b. a.d)$
$d = a + 2b$



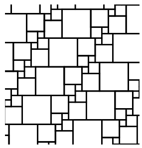
$(d.b.c.d,$
$a.d.d.c.c,$
$a.b.c.b.d.d,$
$a.d.a.c.d.c.b.d.b)$
$d = a + 2b$



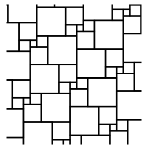
$(d.b.c.d,$
$a.d.c.c.c,$
$a.b.c.b.c.b.d,$
$a.c.b.a.d.a.c.b.$
$a.d)$
$d = a + 2b$



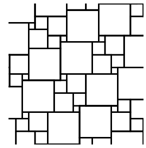
$(c.b.c.b,$
$a.c.d.c.c,$
$a.b.d.d.b.c.b,$
$b.c.d.c.b.c.d.c)$
$d = a + 2b$



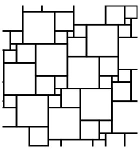
$(b.b.c.c,$
$a.b.a.c.d.c,$
$a.b.d.d.b.a.c,$
$b.c.d.c.b.c.d.c)$
$d = a + 2b$



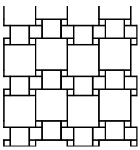
$(b.c.b.c,$
$a.c.d.d.c,$
$a.b.d.d.d.b,$
$b.c.d.c.d.c.b.d)$
$d = a + 2b$



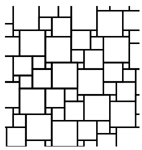
$(b.c.d.d,$
$a.d.d.c.c,$
$a.b.c.b.d.d,$
$a.c.d.c.b.d.a.$
$b.d)$
$b = 2a$
$d = a + 2b$



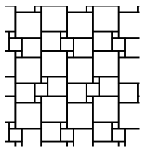
$(c.b.c.b,$
$a.c.d.d.c,$
$a.b.d.d.d.b,$
$b.c.d.c.d.b.c.d)$
$d = a + 2b$



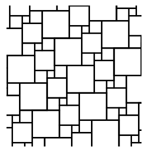
$(d.b.c.d,$
$a.d.c.a.d.c,$
$a.b.d.b.a.d,$
$a.c.a.d.a.b.c.b.$
$a.d)$
$d = a + 2b$



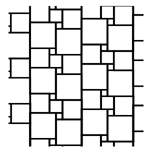
$(d.b.c.d,$
$a.d.d.c.c,$
$a.b.c.b.d.c.d,$
$a.c.c.b.d.a.b.d)$
$d = 2a$
$d = 2a + b$



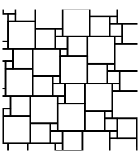
$(d.b.c.d,$
$a.d.c.a.d.c,$
$a.b.d.c.d.d,$
$a.d.c.a.d.c.c.b)$
$b = 2a$
$d = 2a + b$



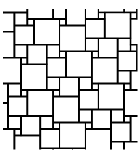
$(c.b.c.b,$
$a.c.d.d.c,$
$a.b.d.b.a.b.d.b,$
$b.c.b.d.b.c.b.d)$
$d = 2a + b$



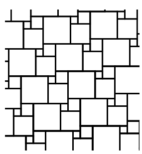
$(d.b.c.d,$
$a.d.c.a.d.c,$
$a.b.d.c.d.d,$
$a.c.d.c.c.b.a.d)$
$b = 2a$
$d = 2a + b$



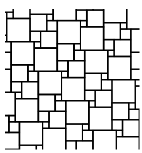
$(d.b.c.d,$
$a.d.c.a.d.c,$
$a.b.d.d.d,$
$a.c.d.c.d.c.b.a.d)$
$d = 2a + b$



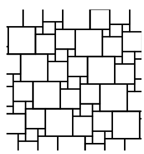
$(d.b.c.d,$
$a.d.d.c.c,$
$a.b.c.d.b.c.d,$
$a.c.c.b.d.a.b.d)$
$b = 2a$
$d = 2a + b$



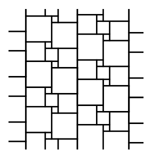
$(b.c.d.c,$
$a.c.d.d.c,$
$a.b.d.b.a.d.d,$
$a.c.d.b.c.b.d.c)$
$d = 2a + b$



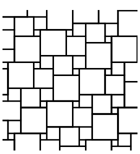
$(d.b.c.d,$
$a.d.d.c.c,$
$a.b.c.b.d.c.d,$
$a.c.c.b.d.b.a.d)$
$b = 2a$
$d = 2a + b$



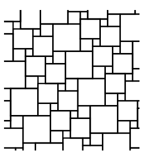
$(c.b.c.d,$
$a.c.c.a.c.c,$
$a.b.c.b.a.d.d,$
$a.c.d.c.a.c.d.c)$
$d = 2a + b$



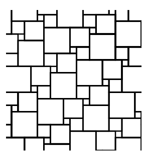
$(b.c.d.d,$
$a.d.c.a.d.c,$
$a.b.d.d.c.d,$
$a.c.c.d.c.b.a.d)$
$d = 2a$
$d = 2a + b$



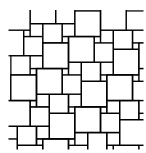
$(b.c.d.d,$
$a.c.c.d.d,$
$a.b.c.d.b.c.d,$
$a.c.c.b.d.a.b.d)$
$d = 2a + b$



$(c.b.c.c,$
$a.c.d.d.c,$
$a.b.d.b.a.c.a.c,$
$b.c.b.d.b.c.b.d)$
$d = 2a + b$



$(c.b.c.d,$
$a.c.d.d.c,$
$a.b.d.a.b.d.d,$
$a.c.d.b.c.d.b.c)$
$d = 2a + b$



$(c.b.c.d,$
$a.c.c.d.d,$
$a.b.c.d.b.c.d,$
$a.c.c.b.d.b.a.d)$
$b = 2a$
$d = 2a + b$

References

[1] G. Grünbaum and G. C. Shephard, *Tilings and Patterns*, Freeman, New York, 1987.

[2] H. Martini, E. Makai and V. Soltan, Unilateral tilings of the plane with squares of three sizes, *Beiträge Algebra Geom.* **39** (1998), 481–495.

[3] D. Schattschneider, Unilateral and Equitransitive Tilings by Squares, *Discrete Comput. Geom.* **24** (2000), 519–526.

ARS MATHEMATICA CONTEMPORANEA

Also available at http://amc-journal.eu
ISSN 1855-3966 (printed edn.), ISSN 1855-3974 (electronic edn.)
ARS MATHEMATICA CONTEMPORANEA 10 (2016) 169–181

Irreducibility of configurations

Klara Stokes *
University of Skövde, School of Engineering Sciences, 54128 Skövde, Sweden

Received 2 February 2014, accepted 24 June 2015, published online 11 September 2015

Abstract

In a paper from 1886, Martinetti enumerated small v_3-configurations. One of his tools was a construction that permits to produce a $(v+1)_3$-configuration from a v_3-configuration. He called configurations that were not constructible in this way irreducible configurations. According to his definition, the irreducible configurations are Pappus' configuration and four infinite families of configurations. In 2005, Boben defined a simpler and more general definition of irreducibility, for which only two v_3-configurations, the Fano plane and Pappus' configuration, remained irreducible. The present article gives a generalization of Boben's reduction for both balanced and unbalanced (v_r, b_k)-configurations, and proves several general results on augmentability and reducibility. Motivation for this work is found, for example, in the counting and enumeration of configurations.

Keywords: Configuration, irreducible, partial linear space, construction, enumeration.

Math. Subj. Class.: 05B30, 51E26, 14N20.

1 Introduction

An incidence geometry is a triple (P, L, I) where P is a set of 'points', L is a set of 'blocks', and I is an incidence relation between the elements in P and L. The line spanned by two points p_1 and p_2 is the intersection of all blocks containing both p_1 and p_2. When there are at most one block containing p_i and p_j for all pairs of points, then we may identify the blocks with the lines. Incidence geometries with this property are called partial linear spaces.

If a point p and a line l are incident, then we say that l goes through p, or that p is on l. We say that a pair of lines that goes through the same point p meet or intersect in p.

*This article was partially written during postdoc positions at Universitat Oberta de Catalunya, Barcelona, Spain and Linköping University, Linköping, Sweden
E-mail address: klara.stokes@his.se (Klara Stokes)

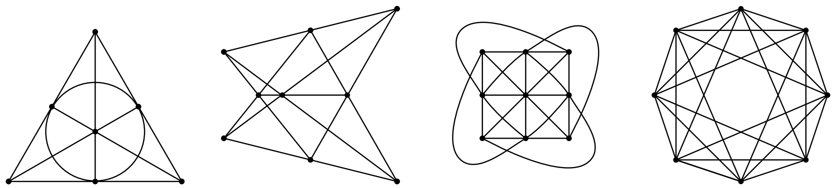


Figure 1: Examples of balanced and unbalanced configurations. From left to right: Fano plane ($v = b = d = 7$ $r = k = 3$), Pappus' configuration ($v = b = d = 7$ $r = k = 3$), Affine plane of order 3 ($v = 9$ $b = 12$ $d = 3$ $r = 4$ $k = 3$), 6-regular graph on 8 vertices ($v = 8$ $b = 24$ $d = 8$ $r = 6$ $k = 2$)

A combinatorial configuration is a partial linear space in which there are r lines through every point and k points on every line [4, 5, 6]. We will use the notation (v_r, b_k)-configuration to refer to a combinatorial configuration with v points, b lines, r lines through every point and k points on every line. The four parameters (v_r, b_k) are redundant so that there is only need for the three parameters (d, r, k), where $d := \frac{v\gcd(r,k)}{k} = \frac{b\gcd(r,k)}{r} = \frac{vr}{\mathrm{lcm}(r,k)} = \frac{bk}{\mathrm{lcm}(r,k)}$ is an integer associated to the configuration that determines the number of points and lines. We will refer to (d, r, k) as the reduced parameter set of the (v_r, b_k)-configuration. When v and b are not known or not important, we will also use the notation (r, k)-configuration.

We say that a configuration is balanced if $r = k$. This implies that the number of points equals the number of lines and the associated integer, so $d = v = b$. In this case, we will use the notation v_k-configuration. In the literature, configurations with this property are also called symmetric. When the configuration is unbalanced, i.e. when $r \neq k$, then v, b and d are all different. Examples of balanced and unbalanced configurations are given in Figure 1.

The following necessary conditions for the existence of configurations are well-known.

Lemma 1.1. *The lower bound of the number of points v of an (r, k)-configuration is $v \geq r(k-1)+1$, and the lower bound of the number of lines b is $b \geq k(r-1)+1$. Also, the parameters v, b, r, k always satisfy $vr = bk$.*

We say that a parameter set satisfying these two conditions are admissible. In general it is difficult to, given some admissible parameter set, determine if there exists some combinatorial configuration with these parameters. If this is the case, then we say that the parameter set is configurable. The (point) deficiency of a configuration with parameters (v_r, b_k) is the difference $\delta_p = v - [r(k-1)+1]$, and the line deficiency is the difference $\delta_l = b - [k(r-1)+1]$. In balanced configurations the two deficiencies are equal.

In 1886 Martinetti studied the construction of v_3-configurations through the addition of a point and a line to existing v_3-configurations [7, 6]. The construction is as follows. Start with a v_3-configuration and assume that there are two parallel lines $\{a, b, c\}$ and $\{a', b', c'\}$ such that a and a' are not collinear. Add a point p and replace the two parallel lines with the lines $\{p, b, c\}$, $\{p, b', c'\}$, $\{p, a, a'\}$. The result is a $(v+1)_3$-configuration. This construction is illustrated in Figure 2. We call such a construction a (Martinetti) augmentation. The inverse construction gives the smaller configuration from the larger one through the re-

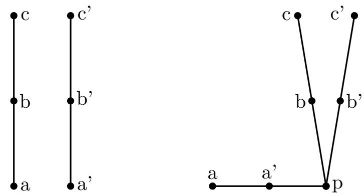


Figure 2: Martinetti's augmentation construction. To the left the two parallel lines in the original v_3-configuration, to the right the new incidences in the constructed $(v+1)_3$-configuration.

moval of one point and one line. We call the inverse construction a (Martinetti) reduction. Martinetti called a configuration irreducible if it could not be constructed from another configuration through an augmentation. In other words, a configuration is irreducible if it does not allow a reduction.

In Martinetti's original paper he gave two infinite families of irreducible v_3-configurations. One consisted of the cyclic configurations with base line $\{0, 1, 3\}$, starting with the smallest v_3-configuration, the Fano plane. There is therefore at least one irreducible v_3 configuration for each $v \geq 7$. The other family gives one irreducible $(10n)_3$-configuration for each $n \in \mathbb{Z}$, starting with Desargues' configuration. Martinetti claimed that these families of configurations were the only irreducible v_3-configurations, with the addition of three sporadic examples for $v \leq 10$; more precisely, Pappus' (9_3)-configuration and two other 10_3-configurations. In 2007, Boben published a correction of this list, in which the two sporadic irreducible 10_3-configurations were shown to be the first elements in two additional infinite families of irreducible $(10n)_3$-configurations, showing that there are four infinite families of irreducible v_3-configurations [2].

Theorem 1.2 (Martinetti - Boben)**.** *The list of (Martinetti) irreducible configurations are*

- *the cyclic configurations with base line* $\{0, 1, 3\}$*. The smallest configuration in this family is the Fano plane,*
- *the three infinite families* $T_1(n)$*,* $T_2(n)$*,* $T_3(n)$*, on* $10n$ *points. The smallest configuration in* $T_1(n)$ *is Desargues' configuration, and*
- *Pappus' configuration.*

It results that, of several possible constructions of $(v+1)_3$ configurations from v_3-configurations, Martinetti's construction is just one example. In 2000, Carstens et al. presented a rather complex set of reductions for which they claimed that the only irreducible configuration was the smallest v_3 configuration - the Fano plane [3]. However, in 2003, Ravnik used computer calculations to show that (at least) the Desargues configuration is also irreducible with respect to this set of reductions [8]. In 2005, Boben presented a simpler definition of reduction in terms of the Levi graph of the configuration. The Levi graph is a lossless representation of the incidences of the points and lines in form of a bipartite graph of girth at least six, and if $r = k = 3$, then it is a cubic graph. In [1], a reduction by the point p and the line l of the v_3-configuration with Levi graph G_v is defined as the Levi graph G_{v-1} of a $(v-1)_3$-configuration obtained from G_v by removing the point vertex p

and the line vertex l from G_v and then connecting their neighbors in such way that the result remains cubic and bipartite. We call this construction a (Boben) reduction. A configuration is (Boben) irreducible if it does not admit a (Boben) reduction. According to Boben, with respect to this reduction, there are only two irreducible v_3-configurations.

Theorem 1.3 (Boben)**.** *The only (Boben) irreducible v_3-configurations are the Fano plane and the Pappus configuration.*

This article presents a generalization of Boben's reduction to (r, k)-configurations for any $r, k \geq 2$, elaborates on the augmentation of v_3 and v_4-configurations and provides some results that ensure irreducibility or reducibility in the general case. Augmentation and reductions of configurations are particularly interesting for the purpose of counting configurations.

2 Reducibility of balanced configurations

Balanced configurations are better studied than unbalanced configurations. This section presents results on augmentation and reduction constructions for balanced configurations.

2.1 Augmentation of balanced configurations

The construction presented next is an augmenting construction for balanced v_k-configurations.

Definition 1. Let $C_v = (P, L, I)$ be a v_k-configuration. Assume that there is a subset of k points $Q \subseteq P$ and a subset of k lines $M \subseteq L$, such that

- there is a bijection $f : Q \to M$ such that the image of a point q is a line $f(q)$ through that point,
- two points $q, q' \in Q$ either are not collinear, or are collinear only on the line $f(q)$ or on the line $f(q')$, and
- two lines $m, m' \in M$ either do not meet, or meet only in the point $f^{-1}(m)$ or in the point $f^{-1}(m')$.

Then there is a $(v + 1)_k$-configuration C_{v+1}, constructed from C_v through the following augmentation procedure:
For all q in Q, replace each incidence $(q, f(q))$ with

- the incidence $(p, f(q))$, where p is a new point, and
- the incidence (q, l), where l is a new line.

Proposition 2.1. *The result of the above construction is a $(v + 1)_k$-configuration C_{v+1}.*

Proof. In C_v, two points $q, q' \in Q$ are either not collinear, or collinear on $f(q)$ or $f(q')$. Since the incidences $(q, f(q))$ and $(q', f(q'))$ have been removed in C_{v+1}, it is clear that in C_{v+1}, the points in Q are collinear only once, on the line l. Analogously, in C_v two lines $m, m' \in M$ either do not meet, or meet only in the point $f^{-1}(m)$ or in the point $f^{-1}(m')$. Since the incidences $(m, f^{-1}(m))$ and $(m', f^{-1}(m'))$ have been removed in C_{v+1}, it is clear that in C_{v+1} the lines in M meet only once, in p. This also shows that any point in C_{v+1} is collinear with p at most once, and that any line in C_{v+1} meets l at most once.

Indeed, the points in C_{v+1} that are collinear with p are the points on the lines in M, and since these lines only meet once in C_{v+1}, we see that any point in C_{v+1} is collinear with p at most once. Also, the lines in C_{v+1} that meet l are the lines through the points in Q, and since these points are collinear only once, in l, we see that any line in C_{v+1} meets l at most once. Now, these are the only incidences affected by the construction, and consequently, it is proved that C_{v+1} is a partial linear space with $v+1$ points and $v+1$ lines. Finally, it is clear that there are k points on each line and k lines through every point, so that C_{v+1} is a $(v+1)_k$-configuration. □

Remark 2.2. The observant reader will find that there are other augmentation constructions which cannot be directly realized by following the steps described above. However, if we allow a final swapping of the incidences involved in the construction, then also these constructions may be described using Proposition 2.1. One example of this is Martinetti's augmentation. Consider $Q = \{a, a', a''\}$ and $M = \{\{a, b, c\}, \{a', b', c'\}, \{a'', b'', c''\}\}$, such that $\{a, b, c\}$ and $\{a', b', c'\}$ are parallel lines and a and a' are not collinear, and no restrictions other than those in Proposition 2.1 are put on a'' and $\{a'', b'', c''\}$, and define $f(a) = \{a, b, c\}$, $f(a') = \{a', b', c'\}$ and $f(a'') = \{a'', b'', c''\}$. Replace the ocurrences of the points in Q on the lines in M with incidences to a new point p so that the resulting lines are $\{p, b, c\}$, $\{p, b', c'\}$, $\{p, b'', c''\}$, and put the points in Q on a new line $\{a, a', a''\}$. Now swap the incidences $(p, \{p, b'', c''\}$ and $(a'', \{a, a', a''\})$ to obtain Martinetti's construction. We see that the original line $\{a'', b'', c''\}$ is then left untouched, in consistency with the fact that Martinetti's construction only involved two lines.

Using Proposition 2.1 it is not difficult to prove the following well-known result.

Corollary 2.3. *There is a v_3-configuration for all admissible parameters.*

Proof. Any v_3-configuration is augmentable. Indeed, if the v_3-configuration has a triangle, then its three points $Q = \{q_1, q_2, q_3\}$ and its three lines $M = \{m_1, m_2, m_3\}$ together with the map $f(q_i) = m_i$ satisfy the conditions in Proposition 2.1. For an illustration of the augmentation in this case, see Figure 3. If the configuration has no triangles, then consider a path starting at a point q_1 of three lines l_1, l_2 and l_3, intersecting in two points q_2 and q_3. Then $Q = q_1, q_2, q_3$ and $M = \{m_1, m_2, m_3\}$ satisfy the conditions of Proposition 2.1. Therefore there is a $(v+1)_3$-configuration whenever there is a v_3-configuration. The smallest v_3-configuration is the Fano plane, with $v = 7$, and the result follows. □

When k is larger than 3, the situation is more complex. Indeed, the projective plane of order 3 is a 13_4-configuration which is not augmentable. However, if a v_4-configuration has at least deficiency one, then it is augmentable.

Corollary 2.4. *There is a v_4-configuration for all admissible parameters.*

Proof. Any v_4-configuration of deficiency at least one is augmentable. Indeed, if the deficiency is at least one, then there are points $Q = \{q_1, q_2, q_3, q_4\}$ and lines $M = \{m_1, m_2, m_3, m_4\}$ forming either a quadrangle with M as sides and Q as vertices, or an open path $q_1 m_1 q_2 m_2 q_3 m_3 q_4 m_4$ such that the conditions of Proposition 2.1 are satisfied. Therefore there is a $(v+1)_4$-configuration whenever there is a v_4-configuration. The smallest v_4-configuration is the projective plane of order 3, and there exists also a 14_4-configuration. This latter configuration has deficiency one, and the result follows. □

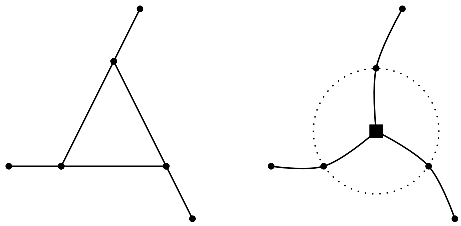


Figure 3: The new (squared) point and the new (plotted) line of a $(v+1)_3$-configuration to the right, added to a triangle in the original v_3-configuration to the left.

2.2 Reduction of balanced configurations

The inverse of the augmentation construction is the reduction.

Definition 2. A reduction of a balanced configuration (P, L, I) is a triple (p, l, f') where p is a point, l is a line, and f' is an injective function $f' : Q' \to M'$, where

- $Q' = \{q \in P : q \in l \text{ and } q \neq p\}$, and
- $M' = \{m \in L : p \in m \text{ and } m \neq l\}$,

such that q is not collinear with $r \in f'(q)$, except possibly through l or with p. A configuration is reducible if it admits a reduction. Otherwise it is irreducible.

Lemma 2.5. *If a configuration (P, L, I) admits a reduction as in Definition 2, then there is a reduced configuration $(P \setminus \{p\}, L \setminus \{l\}, \tilde{I})$ obtained from (P, L, I) by replacing the incidences $(p, f'(q))$ and (q, l) for $q \in Q'$ with the incidences $(q, f'(q))$ and removing the point p and the line l.*

Proof. Each point is on the same number of lines, and each line goes through the same number of points in $(P \setminus \{p\}, L \setminus \{l\}, \tilde{I})$ as in (P, L, I). The definition of f' ensures that any two lines in $(P \setminus \{p\}, L \setminus \{l\}, \tilde{I})$ meet in at most one point. □

Lemma 2.6. *The reduction is the inverse construction of the augmentation.*

Proof. Let $C_v = (P, L, I)$ be a v_k-configuration with a set $Q = \{q_1, \ldots, q_k\}$ of k points and a set $M = \{m_1, \ldots, m_k\}$ of k lines satisfying the requirements in Proposition 2.1. Consider the incidences in the augmented $(v+1)_k$-configuration C_{v+1} which are not in C_v. These incidences are $(p, f(q_i))$ and (q_i, l), for $i \in \{1 \ldots k\}$. Also consider the incidences that were removed from C_v in the construction of C_{v+1}, $(q_i, f(q_i))$, for $i \in \{1 \ldots k\}$. As described in Remark 2.2 and Remark 2.8, some of the incidences involved in the augmentation may be swapped afterwards. This is only relevant if the incidences is of the form $(p, f(q_i))$ and (q_i, l) (which produces the incidence (p, l), so that the new point and the new line are incident). In this case, let $Q' = Q \setminus \{q_i\}$ and $M' = M \setminus \{f(q_i)\}$, otherwise, let $Q' = Q$ and $M = M'$. Define the reduction (p, l, f') with $f' : Q' \to M'$ the restriction of f to Q'. This is a well-defined reduction, since $q \in Q'$ is not collinear with any point r on $f(q)$ in C_{v+1} except possibly with p or through l. Replace the incidences $(p, f(q))$ and (q, l) for $q \in Q'$ with the incidences $(q, f(q))$ and remove the point p and the line l. This reduction produces a v_k-configuration with the same incidences as C_v, hence equal to C_v. □

Observe that according to Definition 2, a balanced configuration is irreducible exactly if it is impossible to remove one point and one line and obtain a new configuration, through modifications that only affect the incidences of the removed point and line. This is the same definition of irreducibility as the one used by Boben in the case of v_3-configurations, although he expressed it in terms of the Levi graph. Martinetti's irreducibility is the special case in which the removed point p is on the removed line $l = \{p, a, a'\}$, so that $Q' = \{a, a'\}$, $M' = \{\{p, b, c\}, \{p, b', c'\}\}$ and $f' : Q' \to M'$ is defined by $f'(a) = \{p, b, c\}$ and $f'(a') = \{p, b', c'\}$. The reduction then consists in removing p and l and replacing the appearances of p in $m \in M$ with $f'^{-1}(m)$. Note that no incidence swapping was needed when describing Martinetti's reduction in terms of Definition 2.

The somewhat awkward definition of reducibility for balanced configurations can also be restated as follows.

Corollary 2.7. *A balanced configuration v_k is reducible if and only if it contains one line l and one point p, such that the points q_i on l and the lines m_i through p can be labelled so that q_i is not collinear with any point on m_i except possibly through l or with p, for $i \in [1, k]$.*

Proof. Indeed, the function $f(p_i) = l_i$ for $p_i \neq p$ gives a reduction (p, l, f). □

Remark 2.8. The general form of the augmentation and reduction constructions implies that the resulting configuration may fail to be connected. However, there is choice in the constructions. It is always possible to make the resulting configuration connected. In practice, this can be achieved by swapping two incidences located in different connected components, as described for example in [9]. That is, if (p, q) and (p', q') are two incidences in two different connected components, then replace these incidences with (p, q') and (p', q). By repeating this process as long as the configuration have at least two connected components, eventually a connected configuration is obtained. If the incidences (p, q) and (p', q') are not incidences of the old configuration, but instead both come from the augmentation or the reduction construction, then the incidence swapping gives a configuration that would have resulted from another choice in the construction. Note that Martinetti's augmentation is described in this way in Remark 2.2.

3 Unbalanced configurations

It is not possible to reduce unbalanced configurations through the removal of one point and one line. This is a consequence of the necessary condition for the existence of a configuration $vr = bk$. Indeed, $vr = bk$ implies that $(v - 1)r/k = vr/k - r/k = b - r/k$ so that $(v - 1)r \neq (b - 1)k$, whenever $r \neq k$. In this context, the reduced parameter set (d, r, k) is useful - the parameter set (d, r, k) is admissible for every integer d satisfying $d \geq \gcd(r, k)(r(k - 1) + 1)/k$. Therefore, a reduction should, given a (d, r, k)-configuration, produce a $(d - 1, r, k)$-configuration through the removal of an appropiate number of points and lines, using only modifications that affect the incidences of these removed points and lines. More precisely, the number of points to remove is $k/\gcd(r, k)$ and the number of lines is $r/\gcd(r, k)$.

3.1 Augmentation of unbalanced configurations

In [9] we described a construction of a $(d_1 + \cdots + d_n + 1, r, k)$-configuration from n configurations with parameters $(d_1, r, k), \ldots, (d_n, r, k)$. By applying this construction to a single configuration with parameters (d, r, k), one obtains a $(d+1, r, k)$-configuration through an augmentation construction. The requirement for this construction to work is that the original configuration contains a set of $rk/\gcd(r,k)$ points Q and a set of $rk/\gcd(r,k)$ lines M with a special property.

Definition 3. Let $C_d = (P, L, I)$ be a (d, r, k)-configuration. Assume that there is a multiset Q of $rk/\gcd(r,k)$ (not necessarily distinct) points in P and a multiset M of $rk/\gcd(r,k)$ (not necessarily distinct) lines in L such that

- there is a bijection $f : Q \to M$ such that the image of a point q is a line $f(q)$ through that point,
- Q can be partitioned into $r/\gcd(r,k)$ parts, each of cardinality k, such that two points q and q' in each part, either are not collinear, or are collinear only on the line $f(q)$ or on the line $f(q')$, and
- M can be partitioned into $k/\gcd(r,k)$ parts, each of cardinality r, such that two lines m and m' in each part either do not meet, or meet only in the point $f^{-1}(m)$ or in the point $f^{-1}(m')$.

Then there is a $(d+1, r, k)$-configuration, constructed from C_d through the following augmentation procedure:
For all q in Q, replace each incidence $(q, f(q))$ with

- the incidence $(p, f(q))$, where p is a point from a set R of $k/\gcd(r,k)$ new points, in a way that ensures that each point in N is on exactly r lines, and
- the incidence (q, l), where l is a line from a set N of $r/\gcd(r,k)$ new lines, in a way that ensures that each line in N contains exactly k points.

Proposition 3.1. *The result of the above construction is a* $(d+1, r, k)$*-configuration.*

The proof of Proposition 3.1 is only slightly more involved than the proof of Proposition 2.1, which is the special case $r = k$. For more details of in the general case, see [9].

Example 3.2. The finite affine plane of order 3 is a $(3, 4, 3)$-configuration (P, L, I) with 9 points and 12 lines (see Figure 1). Label the points P as $1, \dots, 9$ so that the lines L are

$$\{1,2,3\}, \{4,5,6\}, \{7,8,9\}, \{1,4,7\}, \{2,5,8\}, \{3,6,9\},$$

$$\{2,4,9\}, \{3,5,7\}, \{1,6,8\}, \{3,4,8\}, \{1,5,9\}, \{2,6,7\}.$$

An augmentation requires 12 points and 12 lines, and we use $M = L$, Q the multiset consisting of P with the three points $1, 2, 9$ repeated, and the bijection $f : Q \to M$ defined by

$$\begin{array}{llllll} f(1_1) & = \{1,2,3\} & f(1_2) & = \{1,6,8\} & f(2_1) & = \{2,4,9\} \\ f(2_2) & = \{2,6,7\} & f(3) & = \{3,5,7\} & f(4) & = \{3,4,8\} \\ f(5) & = \{1,5,9\} & f(6) & = \{4,5,6\} & f(7) & = \{1,4,7\} \\ f(8) & = \{2,5,8\} & f(9_1) & = \{3,6,9\} & f(9_2) & = \{7,8,9\} \end{array}$$

where x_1 and x_2 denotes the first and the second occurrence of x in Q. This gives, with the new points p_1, p_2, p_3 and the new lines l_1, l_2, l_3, l_4, a $(4, 4, 3)$-configuration with 12 points

$1, \dots, 9, p_1, p_2, p_3$ and 16 lines

$$\begin{array}{llll} \{p_1,1,4\} & \{p_1,2,3\} & \{p_1,5,7\} & \{p_1,6,8\} \\ \{p_2,2,5\} & \{p_2,6,7\} & \{p_2,3,8\} & \{p_2,4,9\} \\ \{p_3,3,6\} & \{p_3,4,5\} & \{p_3,1,9\} & \{p_3,7,8\} \\ \{1,3,7\}=l_1 & \{2,4,8\}=l_2 & \{5,6,9\}=l_3 & \{1,2,9\}=l_4. \end{array}$$

The partition of Q was

$$\{1,3,7\},\{2,4,8\},\{5,6,9\},\{1,2,9\}$$

and the partition of M was

$$\{\{1,2,3\},\{1,4,7\},\{3,5,7\},\{1,6,8\}\},$$

$$\{\{2,4,9\},\{2,5,8\},\{3,4,8\},\{2,6,7\}\},$$

$$\{\{3,6,9\},\{4,5,6\},\{1,5,9\},\{7,8,9\}\}.$$

3.2 Reduction of unbalanced configurations

The inverse of the augmentation construction is a reduction, which is a generalization of the reduction described in Definition 2.

Definition 4. A reduction of an unbalanced configuration (P, L, I) is a triple (R, N, f') where R is a set of $k/\gcd(r,k)$ points, N is a set of $r/\gcd(r,k)$ lines, and f' is a bijection between multisets $f' : Q' \to M'$, where

- $Q' = \{q \in P : \exists l \in N : q \in l \text{ and } q \notin R\}$, and
- $M' = \{m \in L : \exists p \in R : p \in m \text{ and } m \notin N\}$,

such that q is not collinear with $r \in f(q)$, except possibly through one of the lines in N or with one of the points in R. Both Q' and M' are multisets and as such they may contain some element more than once. A configuration is reducible if it admits a reduction. Otherwise it is irreducible.

Lemma 3.3. *If a configuration (P, L, I) admits a reduction as in Definition 4, then there is a reduced configuration $(P \setminus R, L \setminus N, \tilde{I})$ obtained from (P, L, I) by replacing the incidences $(p, f'(q))$ and (q, l) for $q \in Q'$ with the incidences $(q, f'(q))$ and removing the points in R and the lines in N.*

Proof. Each point is on the same number of lines, and each line goes through the same number of points in $(P \setminus \{p\}, L \setminus \{l\}, \tilde{I})$ as in (P, L, I). The definition of f' ensures that any two lines in $(P \setminus R, L \setminus N, \tilde{I})$ meet in at most one point. □

Lemma 3.4. *The reduction is the inverse construction of the augmentation.*

Proof. Let C_d be a (d, r, k)-configuration with a set $Q = \{q_1, \dots, q_{rk/\gcd(r,k)}\}$ of points, a set $M = \{m_1, \dots, m_{rk/\gcd(r,k)}\}$ of lines and a bijection $f : Q \to M$, satisfying the requirements of Definition 3. Consider the incidences in the augmented $(d+1, r, k)$-configuration C_{d+1} which are not in C_d. These incidences are $(p, f(q_i))$ and (q_i, l), for $i \in \{1 \dots rk/\gcd(r,k)\}$, for some $p \in R$ and some $l \in N$ (if no swapping is allowed).

Also consider the incidences that were removed from C_d in the construction of C_{d+1}, $(q_i, f(q_i))$, for $i \in \{1 \dots rk/\gcd(r,k)\}$. If we allow, for some set of indices I, the incidences (q_i, l) and $(p, f(q_i))$, $i \in I$, to be swapped afterwards, making the lines in N and the points in R incident, then let $Q' = Q \setminus \{q_i : i \in I\}$ and $M' = M \setminus \{f(q_i) \in I\}$, otherwise, let $Q' = Q$ and $M = M'$. Define the reduction (R, N, f') with $f' : Q' \to M'$ the restriction of f to Q'. This is a well-defined reduction, since $q \in Q'$ is not collinear with any point r on $f(q)$ in C_{d+1} except possibly with some $p \in R$ or through some $l \in N$. For all $p \in R$ and all $l \in N$, replace the incidences $(p, f(q))$ and (q, l) for $q \in Q'$ with the incidences $(q, f(q))$ and remove the point p and the line l. This reduction produces a (d, r, k)-configuration with the same incidences as C_d, hence equal to C_d. □

Remark 3.5. Remark 2.8, regarding the connectedness of the result of the augmentation and the reduction constructions, is valid also for unbalanced configurations.

4 Irreducibility and reducibility in configurations

We would like to characterize the set of irreducible configurations. The results presented next provide some progress in this direction.

4.1 Irreducibility in small configurations

The smallest (r, k)-configurations are the linear spaces, whenever they exist. Examples of linear spaces are projective and affine planes. The inexistence of smaller (r, k)-configurations clearly implies that the linear spaces are irreducible. However, as the next results states, there are also other (r, k)-configurations that are necessarily irreducible because they are small.

Lemma 4.1. *Any (r, k)-configuration with point deficiency $\delta_p < k - (r+k)/\gcd(r,k)$ or line deficiency $\delta_l < r - (r+k)/\gcd(r,k)$ is irreducible.*

Proof. In a reducible configuration there are points Q' and lines M' and a bijection $f' : Q' \to M'$ such that $q \in Q'$ is not collinear with any of the k points on $f'(q) \in M'$, except possibly with some of the $k/\gcd(r,k)$ removed points R, or through some of the $r/\gcd(r,k)$ removed lines N. This condition is equivalent to requiring that $f'(q) \in M'$ does not meet any of the r lines through q, except possibly on some of the $k/\gcd(r,k)$ removed points R, or through some of the $r/\gcd(r,k)$ removed lines N. But, if the point deficiency $v - [r(k-1)+1]$ is smaller than $k - (r+k)/\gcd(r,k)$, then for any point q there is no line m such that q is only collinear with the points on m on either some points in R or through some lines in N, so the configuration must be irreducible. Analogously, if the line deficiency $b - [k(r-1)+1]$ is smaller than $r - (r+k)/\gcd(r,k)$, then for any line m there is no point $q = f'^{-1}(m)$ such that m does not meet any of the points through q, except if it is a line in N or if the intersection point is a point in R, and again, the configuration must be irreducible. □

This bound is sharp in the meaning that there are reducible (r, k)-configurations of deficiency $\delta_p = k - (r+k)/\gcd(r,k)$ and $\delta_l = r - (r+k)/\gcd(r,k)$. For example, the Möbius-Kantor 8_3-configuration, with deficiency $\delta_p = \delta_l = 3-(3+3)/3 = 1$, is reducible. Indeed, for v_3-configurations, Lemma 4.1 is only relevant for deficiency 0. From [1] we know that the irreducible v_3-configurations are the Fano plane (of deficiency 0) and the Pappus' configuration. However, Pappus' configuration has deficiency 3, so Lemma 4.1

does not apply. But when k is larger than 3, then Lemma 4.1 can imply the irreducibility of more than one (r, k)-configuration. Indeed, for $r = k = 4$, the two configurations 13_4 and 14_4 must be irreducible, for $r = k = 5$, the configurations 21_5 and 23_5 have deficiency 0 and 2, so they are both irreducible. There is no configuration with parameters 22_5. For $r = k = 6$, the configurations 31_6 and 34_6 are both irreducible, since they have deficiencies 0 and 3, hence smaller than 4. There are no configurations 32_6 and 33_6. For $k = 7$, the only configuration with deficiency strictly smaller than 5 that is known to exist is of deficiency 2, with parameters 45_7. There are no configurations 43_7 and 44_7. If the configurations 46_7 and 47_7 exist, then they are irreducible. For a reference on the existence and non-existence of small balanced configurations, see for example [5].

4.2 Irreducible configurations and transversality - Pappus' configuration

The irreducibility of the Fano plane can be explained by Lemma 4.1. The reason why Pappus' configuration is irreducible is different, and based on transversality.

A transversal design $TD_\lambda(k, n)$ is a k-uniform incidence geometry on kn points, allowing a partition of k groups of n elements, such that any group and any block contain exactly one common point, and every pair of points from distinct groups is contained in exactly λ blocks. A transversal design $TD_\lambda(k, n)$ is resolvable if the set of blocks can be partitioned into parallel classes of blocks, such that each class forms a partition of the point set.

When $\lambda = 1$, then the design is a (kn_n, n^2_k)-configuration, and we call the blocks lines. There is a $TD_1(k, n)$ whenever there is an affine plane of order n and $k \leq n$. Indeed, just take the points on k lines in a parallel class of the affine plane and restrict the rest of the lines to these points. Pappus' configuration can be constructed in this way from the affine plane of order 3, by restricting to the points on all the 3 lines in one of its 4 classes of parallel lines. Since the points on these 3 lines are all points in the affine plane, in this case the construction consists of eliminating one parallel class of lines from the affine plane. By instead restricting to the points on only two lines in one of the parallel classes, a transversal design $TD_1(2, 3)$ is obtained, which is a $(6_3, 9_2)$-configuration, that is, the bipartite complete graph on 6 vertices.

Lemma 4.2. *A resolvable transversal design $TD_1(k, n)$ is irreducible if*

$$k \geq (k + r)/\gcd(r, k) + 1.$$

Proof. Let $T = TD_1(k, n)$ be a resolvable transversal design. Let p be a point in T and $m_1, \ldots, m_r$ the lines through p. Then $m_1, \ldots, m_r$ are in different parallel classes. Let l be a line in T and q a point on l. Then q is collinear with all points on the lines $m_1, \ldots, m_r$ except one on each line, which belong to the same group as q (q is not collinear with itself). At most $(r + k)/\gcd(r, k)$ of these incidences will not obstruct a reduction, since a reduction removes $k/\gcd(r, k)$ points and $r/\gcd(r, k)$ lines. Therefore, since the point p and the line l were chosen arbitrarily, if $k \geq (k + r)/\gcd(r, k) + 1$, then it is not possible to find a reduction of T that removes p and l (and possibly other points and lines). More precisely, there is no f' mapping q to m_i, for some i, such that q is not collinear with any point on m_i, except possibly with the $k/\gcd(r, k)$ removed points or through the $r/\gcd(r, k)$ removed lines.

□

Note that this proves that Pappus' configuration, which is a $TD_1(3,3)$, is irreducible, but it does not prove the same fact for the $TD_1(2,3)$. Indeed, the latter is reducible, as is any graph with deficiency high enough. Observe that the deficiency of a transversal design $TD_1(k,n)$ satisfies $d = n-1 \geq k-1$, so that these irreducible configurations are not covered by Lemma 4.1.

4.3 Reducibility in large configurations

When the deficiency is large enough, then reducibility can be ensured.

Lemma 4.3. *A (v_r, b_k)-configuration is reducible if $b \geq 1 + r + r(k-1)(r-1) + r(r-1)^2(k-1)^2$*

Proof. Given a point p there are at most $r + r(k-1)(r-1) + r(r-1)^2(k-1)^2$ lines containing at least one point at distance one or two from p. This bound is attained if the configuration is triangle-, quadrangle-, and pentagonal-free. If the configuration contains an additional line l, then l contains only points at distance at least three from p. In other words, the points on l are not collinear with any point that is collinear with p. This implies that the configuration is reducible. □

In a balanced v_k-configuration, the number of lines b equals the number of points v. Therefore, in this case the bound also takes the form $v \geq 1 + k + k(k-1)^2 + k(k-1)^4$. This is not a sharp bound, indeed, for v_3-configuration it can only ensure reducibility for $v \geq 64$, but we know that all v_3-configurations are reducible for $v \geq 10$.

The irreducibility of v_k-configurations with v between these lower and upper bounds, is still in general an open question. It is of course possible to test a given configuration, by hand or with the help of a computer. However, for exact enumeration purposes it is of course interesting to have exact general results.

5 Conclusions

We have seen that it is possible to define irreducibility not only for (v_k) configuration, but for (v_r, b_k)-configurations in general. Augmentation and reduction constructions for (v_r, b_k)-configurations have been defined in a general manner, and several general results on augmentability and reducibility have been described. Irreducibility has been proved for configurations with point deficiency $\delta_p < k - (r+k)/\gcd(r,k)$ or line deficiency $\delta_l < r - (r+k)/\gcd(r,k)$, and for (some) transversal designs $TD_1(k,n)$. A $TD_1(k,n)$ has point deficiency $n-1 = r-1$ and line deficiency $r^2 - rk + k - 1$. For $r = k = 3$, these are the only irreducible configurations, and at this point, no other irreducible configurations are known in the general case. There is an upper bound for irreducibility requiring the number of lines to satisfy $b < 1 + r + r(k-1)(r-1) + r(r-1)^2(k-1)^2$. This bound is not sharp, and a better bound would probably set the point deficiency closer to r.

The author is aware of at least two applications of augmentations and reductions of configurations. One is the enumeration of configurations, the other is the use of configurations in cryptography and coding theory. When a configuration is used to define a key-distribution scheme, and new parties join or leave, augmentation and reduction constructions can modify the structure while minimizing the costs of key-reassignment. However, it is important to be aware of the fact that the constructions described in this paper may fail to preserve interesting properties.

Acknowledgements

The author would like to thank Mark Ellingham for pointing out that not only the augmentation but also the reduction can result in a disconnected configuration. Partial financial support is acknowledged from the Spanish MEC project ICWT (TIN2012-32757) and ARES (CONSOLIDER INGENIO 2010 CSD2007-00004), as well as from the Swedish National Graduate School in Computer Science (CUGS).

References

[1] M. Boben, (2005) Reductions of (v_3) configurations, arXiv:math/0505136v1.

[2] M. Boben, Irreducible (v_3) configurations and graphs, *Discrete Mathematics* **307** (2007), 331–344.

[3] H. G. Carstens, T. Dinski and E. Steffen, Reduction of symmetric configurations n_3, *Discrete Appl. Math.* **99** (2000), 401–411.

[4] A.A. Davydov, G. Faina, M. Giulietti, S. Marcugini and F. Pambianco, On constructions and parameters of symmetric configurations v_k, *Des. Codes Cryptogr.*, online (2015) 29pp, doi:10.1007/s10623-015-0070-x.

[5] H. Gropp, Configurations, in: C.J. Colbourn, J.H. Dinitz (eds) *The CRC Handbook Of Combinatorial Designs*, Second Edition, CRC Press, Boca Raton, FL, 2007, 352–355.

[6] B. Grünbaum, *Configurations of Points and Lines*, American Mathematical Society, Providence, RI, 2009.

[7] V. Martinetti, Sulle configurazioni piane μ_3, *Annali di mat. pura ed applicata* **15** (1887), 1–26.

[8] E. Steffen, T. Pisanski, M. Boben and N. Ravnik, Erratum to [Reduction of symmetric configurations n_3, Discrete Appl. Math. 99 (1–3) (2000) 401–411], *Discrete Applied Mathematics* **154** (2006), 1645–1646.

[9] K. Stokes and M. Bras-Amorós, Patterns in semigroups associated with combinatorial configurations, in: M. Izquierdo, S.A. Broughton, A.F. Costa, R.E. Rodríguez (eds), *Riemann and Klein Surfaces, Automorphisms, Symmetries and Moduli Spaces, Contemporary Mathematics* **629**, American Mathematical Society, Providence, RI, 2014, 323–333.

Also available at http://amc-journal.eu
ISSN 1855-3966 (printed edn.), ISSN 1855-3974 (electronic edn.)
ARS MATHEMATICA CONTEMPORANEA 10 (2016) 183–192

On $\gamma-$hyperellipticity of graphs

Alexander Mednykh
Sobolev Institute of Mathematics, 630090, Novosibirsk, Russia
Novosibirsk State University, 630090, Novosibirsk, Russia
Siberian Federal University, 660041, Krasnoyarsk, Russia
Univerzita Mateja Bela, 97401, Banská Bystrica, Slovakia

Ilya Mednykh
Sobolev Institute of Mathematics, 630090, Novosibirsk, Russia
Novosibirsk State University, 630090, Novosibirsk, Russia
Siberian Federal University, 660041, Krasnoyarsk, Russia

Received 10 April 2014, accepted 23 July 2015, published online 18 September 2015

Abstract

The basic objects of research in this paper are graphs and their branched coverings. By a graph, we mean a finite connected multigraph. The genus of a graph is defined as the rank of the first homology group. A graph is said to be γ-hyperelliptic if it is a two fold branched covering of a genus γ graph. The corresponding covering involution is called γ-hyperelliptic.

The aim of the paper is to provide a few criteria for the involution τ acting on a graph X of genus g to be γ-hyperelliptic. If τ has at least one fixed point then the first criterium states that there is a basis in the homology group $\mathrm{H}_1(X)$ whose elements are either invertible or split into γ interchangeable pairs under the action of τ_*. The second criterium is given by the formula $\mathrm{tr}_{\mathrm{H}_1(X)}(\tau_*) = 2\gamma - g$. Similar results are also obtained in the case when τ acts fixed point free.

Keywords: Graph, hyperelliptic graph, homology group, Riemann–Hurwitz formula, Schreier formula.

Math. Subj. Class.: 57M12, 57M60

E-mail addresses: smedn@mail.ru (Alexander Mednykh), ilyamednykh@mail.ru (Ilya Mednykh)

1 Introduction

The theory of Riemann surfaces was founded in classical works by B. Riemann and A. Hurwitz. A Riemann surface was originally defined in terms of branched coverings over the Riemann sphere. An important class of surfaces consists of hyperelliptic surfaces, which are defined as branched double coverings of the Riemann sphere.

It is well known that any surface of genus 2 is hyperelliptic. The Farkas-Accola theorem ([1], [6]) states that any unbranched two fold covering of a surface of genus 2 is also a hyperelliptic surface. In [2] Accola showed that an irregular three fold covering of a Riemann surface of genus 2 is also hyperelliptic, while its regular three fold covering is a two fold covering of the torus. Define a Riemann surface be be $\gamma-$hyperelliptic if it is a two fold branched covering of a genus γ surface. The basic properties of γ-hyperelliptic surfaces and their automorphism groups are investigated in the papers ([1], [4], [13]).

Over the last decade, discrete versions of the theory of Riemann surfaces have been addressed in numerous studies. In these theories, finite graphs play the role of Riemann surfaces, while holomorphic mappings are replaced by harmonic ones. Harmonic mappings of graphs are also known as wrapped quasi-coverings or just branched coverings of graphs. The foundation of this theory was done in the paper [12] in terms of dual voltage assignment. Later, the theory was developed from different points of view by H. Urakawa [14], M. Baker and S. Norine [3], S. Corry [5] and others.

The discrete theory found effective applications in coding theory, stochastic theory, and financial mathematics. References concerning this subject can be found in [3].

The basic objects of research in this paper are graphs and their coverings. By a graph, we mean a finite connected multigraph X, possibly with loops.

Denote by $\mathrm{H}_1(X)$ the integer homology group of X. The genus g of a graph X is defined as the rank of $\mathrm{H}_1(X)$ (that is, as the Betti number or the cyclomatic number of a graph). Denote by V and E the number of vertices and edges of X respectively. Then

$$g = 1 - V + E. \tag{1.1}$$

A graph is said to be hyperelliptic if it is a double branched covering of a tree. Note that any two-edge connected graph of genus 2 is hyperelliptic [3].

In this paper we introduce a notion of γ-hyperelliptic graph. A graph is said to be γ-hyperelliptic if it is a two fold branched covering of a genus γ graph. The corresponding covering involution is called $\gamma-$hyperelliptic.

The aim of the paper is to provide some criteria for the involution τ acting on a graph X of genus g to be γ-hyperelliptic. We suppose that τ acts freely and without edge inversion on the set of directed edges of X.

The main results of the paper are Theorems 4.1 and 4.3. Theorem 4.1 gives two criteria for an involution acting on a graph X of genus g with fixed points to be γ-hyperelliptic. Theorem 4.3 provides necessary and sufficient conditions for a γ-hyperelliptic involution to act without fixed points.

2 Preliminary results

In this paper, by a *graph* X we mean a finite connected multigraph, possibly with loops. See, for example, paper [9] for a formal definition of the graph with multiple edges and loops. All edges of X, including loops, are provided by two possible orientations. Denote by $\mathcal{V}(X)$ the set of vertices of X and by $\mathcal{E}(X)$ the set of directed edges of X. We introduce two maps $s,\, t\, :\, \mathcal{E}(X) \to \mathcal{V}(X)$ sending an edge $e \in \mathcal{E}(X)$ to its *source* and *terminate* vertices $s(e)$ and $t(e)$ respectively. We will use also a fixed point free involution $e \to \bar{e}$ of $\mathcal{E}(X)$ (reversal of orientation) such that $s(\bar{e}) = t(e)$ and $s(e) = t(\bar{e})$.

We put $\mathrm{St}(x) = \mathrm{St}^X(x) = \{e \in \mathcal{E}(X) : s(e) = x\}$ for the *star* of x and call $\deg(x) = |\mathrm{St}(x)|$ the *degree* (or *valency*) of x.

2.1 Morphisms of graphs

A morphism of graphs $\varphi : X \to Y$ sends vertices to vertices, edges to edges, and, for any $e \in \mathcal{E}(X),\ \varphi(s(e)) = s(\varphi(e)),\ \varphi(t(e)) = t(\varphi(e)),$ and $\varphi(\bar{e}) = \overline{\varphi(e)}$. For $x \in X$ we then have the local map

$$\varphi_x : \mathrm{St}^X(x) \to \mathrm{St}^Y(\varphi(x)).$$

We call φ *locally surjective or bijective,* respectively, if φ_x has this property for all $x \in X$. We call φ a *covering* if φ is surjective and locally bijective. A bijective morphism is called an *isomorphism*, and an isomorphism $\varphi : X \to X$ is called an *automorphism*.

2.2 Harmonic morphisms

Let $X,\, X'$ be graphs. Let $\varphi : X \to X'$ be a morphism of graphs. We now come to one of the key definitions in this paper.

A morphism $\varphi : X \to X'$ is said to be *harmonic* (or *branched covering*) if, for all $x \in \mathcal{V}(X),\ y \in \mathcal{V}(X')$ such that $y = \varphi(x)$, the quantity $|e \in \mathcal{E}(X) : x = s(e),\, \varphi(e) = e'|$ is the same for all edges $e' \in \mathcal{E}(X')$ such that $y = s(e')$.

We note that an arbitrary covering of graphs is a harmonic morphism. Let $\varphi : X \to X'$ be harmonic and let $x \in \mathcal{V}(X)$. Define the *multiplicity* of φ at x by

$$m_\varphi(x) = |e \in \mathcal{E}(X) : x = s(e),\, \varphi(e) = e'| \tag{2.1}$$

for any edge $e' \in \mathcal{E}(X')$ such that $\varphi(x) = s(e')$. By the definition of a harmonic morphism, $m_\varphi(x)$ is independent of the choice of e'.

If $\deg(x)$ denotes the degree of a vertex x, we have the following formula relating degrees and the multiplicity:

$$\deg(x) = \deg(\varphi(x)) m_\varphi(x). \tag{2.2}$$

We define the degree of a harmonic morphism $\varphi : X \to X'$ by the formula

$$\deg(\varphi) = |e \in \mathcal{E}(X) : \varphi(e) = e'| \tag{2.3}$$

for any edge $e' \in \mathcal{E}(X')$. By the following lemma (see [3], Lemma 2.2) the right-hand side of (2.3) does not depend on the choice of e' (and therefore $\deg(\varphi)$ is well defined):

Lemma 2.1. *The quantity $|e \in \mathcal{E}(X) :\ \varphi(e) = e'|$ is independent of the choice of $e' \in \mathcal{E}(X')$.*

2.3 The γ-hyperelliptic graphs and involutions

A graph X is said to be *γ-hyperelliptic* if there is a two fold harmonic map $\varphi : X \to X'$ on a graph X' of genus γ. That is, a graph is γ-hyperelliptic if it is a two fold branched covering of a genus γ graph. The corresponding covering involution τ is called *γ-hyperelliptic*.

Recall that the covering involution of φ is an order two automorphism τ of graph X satisfying $\varphi \circ \tau = \varphi$.

Since morphism φ is harmonic, by Lemma 2.1 each directed edge of X' has exactly two preimages in the set $\mathcal{E}(X)$ of directed edges of X. Hence, τ permutes these preimages and, consequently, acts freely on the set $\mathcal{E}(X)$. In the category of graphs we deal with, all morphisms send vertices to vertices, edges to edges and loops to loops. In particular, this implies that the covering involution τ acts on X without edge inversion. That is, for every edge e of X we have $\tau(e) \neq \bar{e}$.

On the other hand, if τ is an order two automorphism of a graph X acting freely on the set of directed edges $\mathcal{E}(X)$ and without edge inversion, then the factor space $X' = X/\langle\tau\rangle$ is a graph and the canonical map $X \to X' = X/\langle\tau\rangle$ is harmonic.

Summarizing, we characterize a γ-hyperelliptic involution on graph X as an involution acting on the set $\mathcal{E}(X)$ freely, without edge inversion, and such that genus of the factor graph $X/\langle\tau\rangle$ is γ.

The case $\gamma = 0$ corresponds to hyperelliptic graphs and hyperelliptic involutions defined earlier in [3] in a more general aspect.

2.4 The Riemann-Hurwitz formula for γ-hyperelliptic involution

Let τ be a γ-hyperelliptic involution acting on a graph X of genus g. Then genus of the factor graph $X' = X/\langle\tau\rangle$ is γ. Consider the induced harmonic morphism $\varphi : X \to X'$. By the previous section, τ has neither fixed nor invertible edges, but it may have fixed vertices. For any vertex $a \in \mathcal{V}(X)$ the multiplicity $m_\varphi(x)$ of φ at x is equal to 1 or 2. If $m_\varphi(x) = 1$ then the local map $\varphi_x : \mathrm{St}^X(x) \to \mathrm{St}^{X'}(\varphi(x))$ is bijective. Then the vertex $\varphi(x)$ has two preimages x and $\tau(x)$. So, x is not fixed by τ. If $m_\varphi(x) = 2$ then the local map $\varphi_x : \mathrm{St}^X(x) \to \mathrm{St}^{X'}(\varphi(x))$ is two-to-one on the edges and $\varphi(x)$ has only one preimage $x = \tau(x)$. That is, x is a fixed point of τ.

Denote by V and E the number of vertices and undirected edges of X respectively. Define in a similar way the numbers V' and E' for the graph X'. Since τ acts freely on edges we have $E = 2E'$. Let τ have r fixed points on X. Then there are exactly r vertices of X' with the unique preimage under φ. Hence, $V = 2V' - r$. By formula (1.1) we have $g - 1 = E - V$ and $\gamma - 1 = E' - V'$. Finally, we obtain

$$g - 1 = 2(\gamma - 1) + r. \tag{2.4}$$

This is a discrete version of the classical Riemann-Hurwitz formula from the theory of Riemann surfaces. More general statement of the Riemann-Hurwitz formula for the groups acting on a graph with fixed and invertible edges one can find in [10].

3 Homological basis adapted to the action of an involution

The main results of this section are Theorems 3.2 and 3.5. They can be considered as discrete versions of the results attained earlier for Riemann surfaces by Jane Gilman [7]. We start with the following definition.

Definition 3.1. Let X be a finite connected graph and τ be an involution acting freely and without edge inversion on the set of directed edges of X. Suppose that τ has at least one fixed vertex on X. A homological basis $\mathcal{B}$ in $\mathrm{H}_1(X)$ is said to be *adapted to* τ if it consists of the elements $B_i, C_i, i = 1, \ldots, s,\ D_j, j = 1, \ldots, t$ such that

$$\tau_*(B_i) = C_i,\ \tau_*(C_i) = B_i \text{ and } \tau(D_j) = -D_j.$$

The following theorem yields the conditions on τ to get an adapted homological basis.

Theorem 3.2. *Let X be a finite connected graph and τ is an involution acting freely and without edge inversion on the set of directed edges of X. Suppose that τ has at least one fixed vertex on X. Then $\mathrm{H}_1(X)$ has a basis adopted to τ.*

Proof. We prove the theorem using induction by genus $g = g(X)$. If $g = 0$ then X is a tree, $\mathrm{H}_1(X) = \{0\}$ is a trivial group and $\mathcal{B} = \{0\}$ is a homological basis adapted to τ. For $g > 1$ graph X has at least one nontrivial cycle E. Provide E one of its two possible orientations $\vec{E}$. Let $-\vec{E}$ be the same cycle with the opposite orientation. There are three possibilities:

(i) $\tau(\vec{E}) \neq \pm\vec{E}$,

(ii) $\tau(\vec{E}) = \vec{E}$,

(iii) $\tau(\vec{E}) = -\vec{E}$.

In the case (i) we set $E' = \tau(E)$ and consider an edge $e \in E \setminus E'$. Let $e' = \tau(e)$ then $e' \in E' \setminus E$. Denote by $s(e)$ and $t(e)$ the source and terminate vertices of edge e. Consider the graph $X' = X \setminus \{e, e'\}$. Since e and e' are not in E, the set $E \setminus e$ is a path in X' from $t(e)$ to $s(e)$. In a similar way, $E' \setminus e'$ is a path in X' connecting vertices $t(e')$ and $s(e')$. That is, X' is a connected graph of genus $g(X') = g(X) - 2$. Note that τ leaves the graph X' invariant. Moreover, τ acts without edge inversion on the set of directed edges of X' and has the same set of fixed vertices as before. By induction, X' already has a homological basis $\mathcal{B}'$ adapted to τ. Recall that $\mathcal{B}'$ consist of $g(X')$ elements. Consider the cycles E and E' as elements of $\mathrm{H}_1(X)$. We add E and E' to $\mathcal{B}'$ to form the set $\mathcal{B} = \mathcal{B}' \cup \{E, E'\}$ of homological cycles in $\mathrm{H}_1(X)$. Since edges e and e' are included only in homological cycles E and E' respectively, the set $\mathcal{B}$ consists of $g(X') + 2 = g(X)$ linear independent elements of $\mathrm{H}_1(X)$. Therefore, $\mathcal{B}$ is a τ-adopted basis in $\mathrm{H}_1(X)$.

Now let us consider the case (ii). Since $\tau(\vec{E}) = \vec{E}$, the cycle E is invariant under the action of τ. The involution τ has no fixed points on E since it preserves the orientation and acts freely on the set of directed edges of E. Let v be a fixed point of τ on X, then v is not a vertex of E. Consider a shortest path λ from v to E. Then λ has no common edges with E. Denote by w the terminate point of λ. Set $\lambda' = \tau(\lambda)$ and $w' = \tau(w)$. Then λ' is a path from v to w', and λ' and E have no edges in common. Denote by γ a path in the cycle E from w to w' and by $\gamma' = \tau(\gamma)$ the respective path from w' to w. Consider two cycles $F = \lambda\gamma(\lambda')^{-1}$ and $F = \lambda'\gamma'\lambda^{-1}$. Since the cycle E is not trivial, there exists an edge e in E with the source $s(e) = w$. Then e belongs to F and the respective edge $e' = \tau(e)$ belongs to F'.

Note that $e \in F \setminus F'$ and $e' \in F' \setminus F$. Moreover, graphs $F \setminus \{e\}$ and $F' \setminus \{e'\}$ share v as a common vertex. Hence, the graph $X' = X \setminus \{e, e'\}$ is a connected graph and its genus $g(X') = g(X) - 2$. The involution τ acts on X' satisfying conditions of the theorem. By induction, there is a homological basis $\mathcal{B}'$ in $\mathrm{H}_1(X')$ consisting of $g(X')$ elements and adapted to τ. Consider the set $\mathcal{B} = \mathcal{B}' \cup \{F, F'\}$ of homological cycles in $\mathrm{H}_1(X)$. The elements of $\mathcal{B}$ are linear independent. Indeed, the elements of $\mathcal{B}'$ are already linear independent, but F and F' are the only elements of $\mathcal{B}$ containing edges e and e' respectively. Hence, $g(X)$ elements of $\mathcal{B}$ form a basis in $\mathrm{H}_1(X)$ adapted to τ.

In the case (iii) we have $\tau(\vec{E}) = -\vec{E}$. It means that τ leaves cycle E invariant reversing the orientation of its edges. By assumption, τ acts without edge inversion. Then τ has exactly two fixed points v and w on E. Let e be the edge of E with the source $s(e) = v$ and $e' = \tau(e)$. Consider the graph $X' = X \setminus \{e, e'\}$.

Suppose that graph X' is connected. Then X' contains a path γ from v to w. The union $E \cup \gamma$ is a connected graph of genus at least two. So, it contains a cycle F passing through e which differs from E. Then $\tau(\vec{F}) \neq \vec{F}$ and the proof follows from the case (i).

Now we suppose that graph X' is disconnected. The X' consists of two connected components X'_1 and X'_2 containing the vertices v an w respectively. Since v and w are fixed by τ, both X'_1 and X'_2 satisfy the conditions of the theorem. Also we have $g(X) = g(X'_1) + g(X'_2) + 1$.

By inductive assumption, X'_1 and X'_2 admit homological bases $\mathcal{B}'_1$ and $\mathcal{B}'_2$ adapted to τ. Consider the cycle E as an element of $\mathrm{H}_1(X)$. Then the set $\mathcal{B} = \mathcal{B}'_1 \cup \mathcal{B}'_2 \cup \{E\}$ gives a homological basis in $\mathrm{H}_1(X)$ adapted to τ. □

Remark 3.3. The condition on τ to have fixed points in Theorem 3.2 is essential. Indeed, consider a cyclic graph X on even number of vertices. Let τ act on X by the order two rotation. Then τ acts fixed point free on the vertices of X. The homology group $\mathrm{H}_1(X)$ is generated by a cycle A and $\tau_*(A) = A$. That is, τ_* acts trivially on $\mathrm{H}_1(X)$.

To investigate the action of τ without fixed points, we introduce the following definition.

Definition 3.4. Let X be a finite connected graph and τ be an involution acting on X without fixed vertices. A homological basis $\mathcal{B}$ in $\mathrm{H}_1(X)$ is said to be *adapted to* τ if it consists of the elements $A, B_i, C_i, i = 1, \dots, s$ such that

$$\tau_*(A) = A,\ \tau_*(B_i) = C_i,\ \tau_*(C_i) = B_i,\ i = 1, \dots, s.$$

Theorem 3.5. *Let X be a finite connected graph of genus g. Let τ be an involution acting on X without edge inversion. Suppose that τ has no fixed vertices. Then genus g is an odd number and $\mathrm{H}_1(X)$ has a basis adapted to τ.*

Proof. The first statement of the theorem follows from the Riemann-Hurwitz formula. In our case, it has the form $g - 1 = 2(\gamma - 1)$, where γ is the genus of the factor graph $X/\langle\tau\rangle$. Since $\gamma \geq 0$ we have $g \geq 1$. Hence, $g = 2(\gamma - 1) + 1$ is a positive odd number.

We prove the second statement of the theorem using induction by genus g. If $g = 1$ then X has only one cycle E. Provide E one of its two possible orientations $\vec{E}$. Then $\tau(\vec{E}) = \pm\vec{E}$. By assumption, τ acts without edge inversion. Then in the case $\tau(\vec{E}) = -\vec{E}$ it has two fixed points. This is impossible, since τ acts fixed point free. So, $\tau(\vec{E}) = \vec{E}$ and $\mathcal{B} = \{E\}$ is the basis in $\mathrm{H}_1(X)$ adapted to τ.

Since $g = g(X)$ is an odd number, one can assume that $g \geq 3$. Consider an arbitrary cycle E in X. Choosing an orientation $\vec{E}$ on E we have two possibilities: (i) $\tau(\vec{E}) \neq \pm\vec{E}$ and (ii) $\tau(\vec{E}) = \vec{E}$. By the above arguments, the case $\tau(\vec{E}) = -\vec{E}$ is impossible.

In the case (i) we set $E' = \tau(E)$. Since $E \neq E'$, there are edges $e \in E$ and $e' = \tau(e) \in E'$ such that $e \in E \setminus E'$ and $e' \in E' \setminus E$. Then $X' = X \setminus \{e, e'\}$ is a connected graph of genus $g(X') = g(X) - 2 \geq 1$. Graph X' is invariant under τ and satisfies the conditions of the theorem. By induction, $\mathrm{H}_1(X')$ has a τ-adapted basis $\mathcal{B}'$. Consider the cycles E and E' as elements of $\mathrm{H}_1(X)$. Then $\mathcal{B} = \mathcal{B}' \cup \{E, E'\}$ is a basis in $\mathrm{H}_1(X)$ adopted to τ.

In the case (ii) we have $\tau(E) = E$. Choose an edge $e \in E$ and set $e' = \tau(e)$. Then $e' \in E$ and $e \neq e'$. Consider the graph $X = X' \setminus \{e, e'\}$. If X' is disconnected, it consists of two components X'_1 and X'_2 permuted by τ. Thereby $g(X'_1) = g(X'_2) = g'$ and $g(X) = 2g' + 1$. Let the group $\mathrm{H}_1(X'_1)$ be generated by cycles $B_1, B_2, \ldots, B_{g'}$. Then $\mathrm{H}_1(X'_2)$ is generated by cycles $C_1, C_2, \ldots, C_{g'}$, where $C_i = \tau(B_i), i = 1, \ldots, g'$ and

$$\mathcal{B} = \{B_1,\, B_2, \ldots, B_{g'},\, C_1,\, C_2, \ldots, C_{g'}, E\}$$

is the required τ-adapted basis in $\mathrm{H}_1(X)$.

Now, let the graph $X = X' \setminus \{e, e'\}$ be connected. Then there is a path γ from $s(e)$ to $t(e)$ in X'. The genus of graph $E \cup \gamma$ is at least two, so $E \cup \gamma$ contains a cycle F such that $\tau(F) \neq F$. Then $\tau(\vec{F}) \neq \pm\vec{F}$ and we are back to the case (i). □

4 Main results

Now we apply the Theorems 3.2 and 3.5 to establish the main results of the paper. They are given in Theorems 4.1 and 4.3.

Theorem 4.1 gives two criteria for an involution acting on a graph X of genus g with fixed points to be γ-hyperelliptic. Theorem 4.3 provides necessary and sufficient conditions for a γ-hyperelliptic involution to act without fixed points.

Theorem 4.1. *Let X be a finite connected graph of genus g. Consider an involution τ acting freely and without edge inversion on the set of directed edges of X. Denote by τ_* the induced action of τ on the first homology group $\mathrm{H}_1(X)$. Suppose that τ has at least one fixed vertex on X. Then the following conditions are equivalent:*

- *(i) the genus of factor graph $X/\langle\tau\rangle$ is equal to γ;*
- *(ii) there is a basis in the homology group $\mathrm{H}_1(X)$ whose g elements are either invertible or split into γ interchangeable pairs under the action of τ_*;*
- *(iii) $\mathrm{tr}_{\,\mathrm{H}_1(X)}\,(\tau_*) = 2\gamma - g$.*

Proof. Suppose that τ has $r \geq 1$ fixed points on $\mathcal{V}(X)$.

We show that (i) implies (ii). By Theorem 3.2 there exists a homological basis $\mathcal{B}$ in $\mathrm{H}_1(X)$ adapted to τ. Without loss of generality, we can assume that $\mathcal{B}$ consists of s interchangeable pairs $B_i, C_i,\ i = 1, \ldots, s$, and $D_j, j = 1, \ldots, t$ invertible elements, where s and t are non-negative integers related by the equation

$$2s + t = g. \tag{4.1}$$

Then the induced action τ_* of τ on $\mathrm{H}_1(X)$ is given by the formulas:

$$\tau_*(B_i) = C_i,\ \tau_*(C_i) = B_i,\ i = 1, \ldots, s \text{ and } \tau_*(D_j) = -D_j,\ j = 1, \ldots, t.$$

Consider the matrix representation of τ_* in the basis $\mathcal{B}$. Then we have:

$$\tau_* = \underbrace{\left(\begin{array}{cccccc} J & \dots & 0 & 0 & \dots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \dots & J & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & -1 & \dots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & 0 & \dots & -1 \end{array}\right)}_{2s \quad\quad t} \begin{array}{l} \left.\vphantom{\begin{array}{c}1\\1\\1\end{array}}\right\} 2s \\ \\ \left.\vphantom{\begin{array}{c}1\\1\\1\end{array}}\right\} t \end{array},$$

where

$$J = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Using the above matrix representation, by direct calculation we obtain

$$\operatorname{tr}_{\mathrm{H}_1(X)}(\tau_*) = -t. \tag{4.2}$$

At the same time, by the Hopf-Lefschetz formula [8] we have

$$\operatorname{tr}_{\mathrm{H}_1(X)}(\tau_*) = 1 - r, \tag{4.3}$$

where r is the number of fixed point of τ on the graph X. Now we use the Riemann-Hurwitz formula

$$g - 1 = 2(\gamma - 1) + r \tag{4.4}$$

to find s and t through g and γ. From (4.2), (4.3) and (4.4) we have $t = r - 1 = g - 2\gamma$. Hence, taking into account (4.1) we conclude that $s = \gamma$.

These are the statements (iii) and (ii) of the theorem.

We note that (iii) follows from (ii) by direct calculations. To finish the proof, we have to show that (iii) implies (i). Again, by the Hopf-Lefschetz formula we have (4.3). Hence $2\gamma - g = 1 - r$. Equivalently, $g - 1 = 2(\gamma - 1) + r$. By Riemann-Hurwitz formula we conclude that γ is genus of the factor graph $X/\langle\tau\rangle$. □

The equivalence of conditions (i) and (ii) for $\gamma = 0$ was established earlier by the second named author in ([11], Lemma 1).

The following result is an immediate consequence from the proof of Theorem 4.1.

Corollary 4.2. *Let τ be the same as in Theorem 4.1 and*

$$\{B_i, C_i, i = 1, \dots, \gamma, D_j, j = 1, \dots, t\}, \tau_*(B_i) = C_i,\ \tau_*(C_i) = B_i,\ \tau_*(D_j) = D_j,$$

be a homological basis in $\mathrm{H}_1(X)$ *adapted to* τ. *Then* $g(X/\langle\tau\rangle) = \gamma$ *and the number of fixed vertices of* τ *is* $t + 1$.

Theorem 4.3. *Let X be a finite connected graph of genus $g \geq 1$. Consider an involution τ acting freely and without edge inversion on the set of directed edges of X. Denote by τ_* the induced action of τ on the first homology group $\mathrm{H}_1(X)$ and by γ the genus of factor graph $X/\langle\tau\rangle$. Then τ acts fixed point free on the set of vertices of X if and only if one of the following conditions is satisfied:*

(i) *genera g and γ are related by the Schreier formula $g - 1 = 2(\gamma - 1)$;*

(ii) *there is a basis $\{A, B_i, C_i, i = 1, \ldots, \gamma - 1\}$ in the homology group $\mathrm{H}_1(X)$ such that $\tau_*(A) = A,\ \tau_*(B_i) = C_i,\ \tau_*(C_i) = B_i$;*

(iii) $\mathrm{tr}_{\,\mathrm{H}_1(X)}\,(\tau_*) = 1$.

Proof. Let r be the number of fixed vertices of τ. Then (i) immediately follows from the Riemann-Hurwitz formula (4.4).

We show that (ii) and (i) are equivalent. Indeed, by Theorem 3.5 the group $\mathrm{H}_1(X)$ has a basis $\mathcal{B}$ adapted to τ and consisting of g elements. By (i) we have $g = 2(\gamma - 1) + 1$. So, $\gamma - 1$ elements of $\mathcal{B}$ are permutable by τ_* and one of them is fixed by τ_*. This is exactly the statement (ii).

To prove that (ii) implies (i) we note that the basis $\mathcal{B} = \{A, B_i, C_i, i = 1, \ldots, \gamma - 1\}$ of the $\mathrm{H}_1(X)$ consists of $2(\gamma - 1) + 1$ elements. Hence, $g = 2(\gamma - 1) + 1$ and the Schreier formula holds. Calculating trace of τ_* in the basis $\mathcal{B}$, we obtain $\mathrm{tr}_{\,\mathrm{H}_1(X)}\,(\tau_*) = 1$. Therefore, (iii) follows from (ii). Finally, let $\mathrm{tr}_{\,\mathrm{H}_1}(X)\,(\tau_*) = 1$. Then, by the Hopf-Lefschetz formula the number r of the fixed points of τ is equal to 0. Then again, by (4.4) we get (i). □

5 Acknowledgments

The authors are grateful to the referee for the useful remarks and suggestions that leads to the essential improvement of the paper.

The authors are thankful to the following grants for partial support of this investigation: the Russian Foundation for Basic Research (grant 15-01-07906), the Grant of the Russian Federation Government at Siberian Federal University (grant 14.Y26.31.0006), the Project Mobility–Enhancing Research, Science and Education, Matej Bel University (ITMS code 26110230082) under the Operational Program Education cofinanced by the European Social Foundation, and by the Slovenian-Russian grant (2014-2015).

References

[1] R. D. M. Accola, Riemann surfaces with automorphism groups admitting partitions, *Proc. Amer. Math. Soc.* **21** (1969), 477–482.

[2] R. D. M. Accola, On lifting the hyperelliptic involution, *Proc. Amer. Math. Soc.* **122** no. 2 (1994), 341–347.

[3] M. Baker and S. Norine, Harmonic morphisms and hyperelliptic graphs, *Int. Math. Res. Notes* **15** (2009), 2914–2955.

[4] E. Bujalance and A. F. Costa, On symmetries of $p-$hyperelliptic Riemann surfaces, *Math. Ann.* **308** no. 1 (1997), 31–45.

[5] S. Corry, Genus bounds for harmonic group actions on finite graphs, *Inter. Math. Res. Not.* **19** (2011), 4515–4533.

[6] H. M. Farkas, Unramified double coverings of hyperelliptic surfaces, *J. Anal. Math.* **30** (1976), 150–155.

[7] J. Gilman, A matrix representation of automorphisms of compact Riemann surfaces, *Linear Algebra and Its Applications* **17** (1977), 139–147.

[8] O. Knill, A Brouwer fixed-point theorem for graph endomorphisms, *Fixed Point Theory and Applications* **2013** no.1 (2013), 1–24.

[9] A. Malnič, R. Nedela and M. Škoviera, Lifting graph automorphisms by voltage assignments, *European Journal of Combinatorics* **21** no. 7 (2000), 927–947.

[10] A. D. Mednykh, On the Riemann-Hurwitz formula for graph coverings, (2015), Preprint arXiv:1505.00321.

[11] I. A. Mednykh, Discrete analogs of Farkas and Accolas theorems on hyperelliptic coverings of a Riemann surface of genus 2, *Mathematical Notes* **96** no. 1–2 (2014), 84–94.

[12] T. D. Parsons, T. Pisanski and P. Jackson, Dual imbeddings and wrapped quasi-coverings of graphs, *Discrete Mathematics* **31** no. 1 (1980), 43–52.

[13] E. Tyszkowska, On $p-$hyperelliptic involutions of Riemann surfaces, *Contributions to Algebra and Geometry* **46** no. 2 (2005), 581–586.

[14] H. Urakawa, A discrete analogue of the harmonic morphism and Green kernel comparison theorems, *Glasg. Math. J.* **42** no. 3 (2000), 319–334.

Also available at http://amc-journal.eu
ISSN 1855-3966 (printed edn.), ISSN 1855-3974 (electronic edn.)
ARS MATHEMATICA CONTEMPORANEA 10 (2016) 193–209

Classification of the regular oriented hypermaps with prime number of hyperfaces*

Antonio Breda d'Azevedo , Maria Elisa Fernandes

Center for Research and Development in Mathematics and Applications
Department of Mathematics, University of Aveiro, Aveiro, Portugal.

This paper is part of the Special Issue in Honor of the 60th Birthday of Professor Dragan Marušič

Received 30 April 2014, accepted 6 March 2015, published online 18 September 2015

Abstract

Regular oriented hypermaps are triples $(G; a, b)$ consisting of a finite 2-generated group G and a pair a, b of generators of G, where the left cosets of $\langle a\rangle$, $\langle b\rangle$ and $\langle ab\rangle$ describe respectively the hyperfaces, hypervertices and hyperedges. They generalise regular oriented maps (triples with ab of order 2) and describe cellular embeddings of regular hypergraphs on orientable surfaces. Previously, we have classified the regular oriented hypermaps with a prime number of hyperfaces and with no non-trivial regular proper quotients with the same number of hyperfaces (i.e. primer hypermaps with prime number of hyperfaces), which generalises the classification of regular oriented maps with prime number of faces and underlying simple graph. Now we classify the regular oriented hypermaps with a prime number of hyperfaces. As a result of this classification, we conclude that the regular oriented hypermaps with prime p hyperfaces have metacyclic automorphism groups and the chiral ones have cyclic chirality groups; of these the "canonical metacyclic" (i.e. those for which $\langle a\rangle$ is normal in G) have chirality index a divisor of n (the hyperface valency) and the non "canonical metacyclic" have chirality index p. We end the paper by counting, for each positive integer n and each prime p, the number of regular oriented hypermaps with p hyperfaces of valency n.

Keywords: hypermaps, maps, hypergraphs, regularity, orientably regular, chirality

Math. Subj. Class.: 05E18, 05E15, 20B25, 05C25, 05C30

*This work was supported by Portuguese funds through the CIDMA–Center for Research and Development in Mathematics and Applications, and the Portuguese Foundation for Science and Technology (FCT–Fundação para a Ciência e a Tecnologia), within project PEst-OE/MAT/UI4106/2014.

E-mail addresses: breda@ua.pt (Antonio Breda d'Azevedo), maria.elisa@ua.pt (Maria Elisa Fernandes)

1 Introduction

Hypermaps (surface embeddings of hypergraphs), introduced by Cori [10] in 1975, have acquired great importance in recent years as a connection between permutations, extended triangle groups, Riemann surfaces, algebraic curves and Galois groups. As highlighted by Grothendieck [14], the absolute Galois group of the field of algebraic numbers acts faithfully on dessins d'enfants (hypermaps), combinatorial objects that, by Belyi's theorem [1], characterise the Riemann surfaces defined (as projective algebraic curves) over the field of algebraic numbers. The correspondence between hypermaps and Riemann surfaces is in general difficult to study, but becomes more manageable if the hypermaps are uniform (that is, if all hyperfaces have the same size n, all hypervertices have the same degree k and all hyperedges have the same size m) and particularly better handled when they are regular.

In this paper we concentrate on regular oriented hypermaps, which are algebraically characterised by triples $\mathcal{H} = (G; a, b)$ consisting of a finite 2-generated group G and a pair a, b of generators of G; such triples encode cellular embeddings of regular hypergraphs (bipartite graphs[1]) on compact orientable surfaces of genus

$$g = \frac{2 - (|G/_l\langle b\rangle| + |G/_l\langle ab\rangle| + |G/_l\langle a\rangle|)}{2},$$

where $G/_l H$ stands for the left cosets of the subgroup H in G and $|X|$ the cardinality of X. The left cosets of $\langle a\rangle$, $\langle b\rangle$ and $\langle ab\rangle$ determine the hyperfaces, hypervertices and hyperedges of $\mathcal{H}$. Regular oriented maps are regular oriented hypermaps $(G; a, b)$ in which the product ab has order 2.

Regular (cellular-) embeddings of graphs in orientable surfaces (regular orientable maps) have been classified for certain classes of graphs. The closest to the present paper is the classification of orientable regular embeddings of graphs of given order. This has been achieved for simple graphs of prime order [13] and of order a product of two primes [12], giving rise respectively to classifications of the regular oriented simple maps of prime order, and of order a product of two primes. Regular oriented maps of type $\{|a|, |b|\}$ are regular hypermaps of type $(|b|, 2, |a|)$. Here $|g|$ is the order of g.

Up to a duality a primer hypermap is a generalisation of a simple map (map with underlying simple graph). In [4] we classified the primer hypermaps with prime number of hyperfaces (left cosets of $\langle a\rangle$ in G) and now we extend the classification to regular oriented hypermaps with prime number of hyperfaces – or, by duality, to a classification of the regular oriented hypermaps with prime number of hypervertices (left cosets of $\langle b\rangle$) or hyperedges (left cosets of $\langle ab\rangle$).

There has been some contributions to the classification of regular (oriented or non-oriented) hypermaps by given number of hyperfaces; namely, on regular hypermaps (non-oriented hypermaps, which include non-orientable hypermaps and hypermaps with border) with one and two hyperfaces [9], on non-orientable regular hypermaps with a prime number of hyperfaces [19], on chiral hypermaps up to 4 hyperfaces [7], and on regular oriented hypermaps up to 5 hyperfaces [3].

This paper has five sections. The first is the actual introduction which includes two subsections, one giving a quick overview of the theory of regular oriented hypermaps and the second giving an overview of "primer" hypermaps. In this subsection we write down the most important results of [4] that are used in the third section. For a complementary

[1] Graphs in this paper are pseudographs, that is, they may have multiple edges, loops and free-edges.

reading on these subjects we address the reader to [16, 17, 11, 7, 5, 4]. In the second section we introduce some families of hypermaps, called "derivations", that arise from a given regular oriented hypermap, and explore their properties. The third is the classification of the regular oriented hypermaps with p (prime) hyperfaces, and this will be achieved by "lifting" the "primer" hypermaps with p hyperfaces classified in [4]. In the fourth section, we compute the chirality group and the "H-sequences" (an extension of type) of the regular oriented hypermaps with p hyperfaces. And finally in the fifth section we compute the number of regular oriented hypermaps with p hyperfaces of valency n.

Functions in this paper are read from right to left.

1.1 Regular oriented hypermaps

An (finite) *oriented hypermap* is a triple $\mathcal{H} = (\Omega; a, b)$ consisting of a finite set Ω (the set of *darts*) and two permutations a and b that generate a transitive group G (called the *monodromy* group) on Ω. *Hyperfaces*, *hypervertices* and *hyperedges* of $\mathcal{H}$ are orbits of $\langle a\rangle$, $\langle b\rangle$ and $\langle ab\rangle$ respectively, and incidence is given by non-empty intersection of orbits. Here ab means a followed by b since functions and actions in this paper are supposed to act from right. $\mathcal{H}$ is *uniform* if the permutations a, b and ab are regular permutations; this means that all the hyperfaces have common valency, all the hypervertices have common degree and all the hyperedges have common valency. In general we have $|\Omega| \geq |G|$. If $|\Omega| = |G|$, that is, if G acts regularly on Ω, then we say that $\mathcal{H}$ is a *regular* oriented hypermap. In such case Ω can be replaced by G and the right actions of a and b by right multiplication. Conversely, any finite two generated group $G = \langle a, b\rangle$ determines a regular oriented hypermap $(G; \overline{a}, \overline{b})$ where the monodromy elements $\overline{a}$ and $\overline{b}$ are the respective right permutation representations of a and b on G. The above triple describes a cellular embedding of a hypergraph $\mathcal{G}$ in an oriented surface $\mathcal{S}$ (i.e., an orientable surface with a fixed orientation). Viewing $\mathcal{G}$ as a bipartite graph, with the set of vertices partitioned into black vertices and white vertices, the hypermap $\mathcal{H}$ can be seen as a bipartite map $\mathcal{M}$ where the black vertices of $\mathcal{G}$ represent the hypervertices, the white vertices the hyperedges and the faces of $\mathcal{M}$ the hyperfaces [18]. In this representation the edges of $\mathcal{G}$ are the *darts* of $\mathcal{H}$ and the permutations a and b locally permute the darts counter clockwise (CCW) around hyperfaces and hypervertices respectively (actually in the literature it is more common a and b be permutations of darts CCW around hypervertices and hyperedges, and usually denoted by R and L).

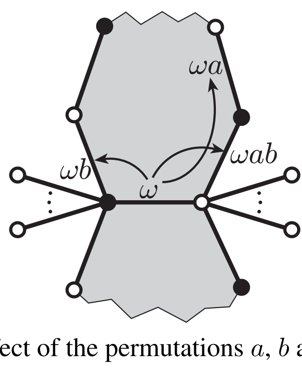


Figure 1: The effect of the permutations a, b and ab on a dart ω.

The *type* of a regular oriented hypermap $\mathcal{H}$ is a triple (k, m, n) where the positive integers $k = |b|$, $m = |ab|$, $n = |a|$ are the valencies of the hypervertices, hyperedges and hyperfaces, in this order. An extended version of the type is the *H-sequence* $[k, m, n\,;\, V, E, F\,;\, |G|]$ where (k, m, n) is the type, V, E and F are respectively the number of hypervertices, hyperedges and hyperfaces, and $|G|$ is the size of G (or the number of darts of $\mathcal{H}$). The Euler characteristic of the underlying surface $\mathcal{S}$ is the *characteristic of* $\mathcal{H}$, and it is given by the formula $\chi = V + E + F - |G|$.

If $\mathcal{H} = (G; a, b)$ and $\mathcal{H}' = (G'; a', b')$ are two regular oriented hypermaps, then $\mathcal{H}$ *covers* H' if the assignment $a \mapsto a'$, $b \mapsto b'$ can be extended to a (*canonical*) epimorphism of monodromy groups $G \to G'$. The hypermap $\mathcal{H}$ is *isomorphic* to $\mathcal{H}'$, $\mathcal{H} \cong \mathcal{H}'$, if the canonical epimorphism $G \mapsto G'$, is an isomorphism. A hypermap is *reflexible* if it is isomorphic to its mirror image $\overline{\mathcal{H}} = (G; a^{-1}, b^{-1})$, otherwise it is *chiral*. The *chirality group* of $\mathcal{H}$ is the smallest normal subgroup $X(\mathcal{H})$ of G such that $\mathcal{H}/X(\mathcal{H})$ is reflexible. This group ranges from $X(\mathcal{H}) = 1$, when $\mathcal{H}$ is reflexible, to $X(\mathcal{H}) = Mon(\mathcal{H})$ when $\mathcal{H}$ is totally chiral [5, 6]. The Chirality index of $\mathcal{H}$ is the size $\kappa = \kappa(\mathcal{H}) = |X(\mathcal{H})|$.

Let Δ denote the free product $C_2 * C_2 * C_2$ generated by r_0, r_1 and r_2, and Γ be the normal subgroup of index 2 in Δ generated by $a = r_0r_1$ and $b = r_1r_2$, a free group of rank 2. Any regular oriented hypermap $\mathcal{H}$ corresponds an unique normal subgroup H in Γ, called the *fundamental hypermap subgroup*, such that $\mathcal{H} \cong (\Gamma/H; Ha, Hb)$. In this context the chirality group of $\mathcal{H}$ is given by $X(\mathcal{H}) = H\overline{H}/H$, where $\overline{H} = H^{r_1}$. If $\langle a, b \mid R(a, b)\rangle$ is a presentation of the monodromy group G, where $R(a, b)$ denotes a set of relators on a and b, then the chirality group of $\mathcal{H}$ is $X(\mathcal{H}) = \langle R(a^{-1}, b^{-1})\rangle^G$, the normal closure in G of the subgroup generated by $R(a^{-1}, b^{-1})$ [2].

The regular oriented hypermaps $\mathcal{H} = (G; a, b)$ with 1 and 2 hyperfaces are all reflexible and the chiral hypermaps with 3 and 4 hyperfaces are all *(face-)canonical metacyclic*, that is, the monodromy group G is the metacyclic group $\langle a, b \mid a^n = 1,\, b^m = a^s,\, bab^{-1} = a^t\rangle$ with $(t - 1)s = 0 \mod n$ and $t^m = 1 \mod n$. Equally we say that $(G; a, b)$ is *vertex-canonical* (resp. *edge-canonical*) metacyclic if $\langle b\rangle$ (resp. $\langle ab\rangle$) is normal in G. $\mathcal{H}$ is vertex-canonical (resp. edge-canonical) metacyclic if and only if $\mathcal{H}\delta_1$ (resp. $\mathcal{H}\delta_0$) is face-canonical metacyclic, where δ_1 is the dual operation $a \mapsto b^{-1}$, $b \mapsto a^{-1}$ that transpose hypervertices with hyperfaces, and δ_0 is the dual operation $a \mapsto ab$, $b \mapsto b^{-1}$ that transpose hyperedges with hyperfaces. Another dual operation is the *mirror* operation $\mu : a \mapsto a^{-1}$, $b \mapsto b^{-1}$ that maps $\mathcal{H}$ to its mirror image $\mathcal{H}\mu = \overline{\mathcal{H}}$. Face-, vertex- and edge-canonical metacyclic hypermaps have cyclic chirality groups with chirality index $\frac{n}{\gcd(n,\, t^2-1)}$; while the chirality group of a face-canonical hypermap is generated by a^{t^2-1} [7], the chirality group of a vertex- or edge-canonical metacyclic hypermap is generated by $(a^{t^2-1}\delta_1)^{-1} = b^{t^2-1}$ or by $a^{t^2-1}\delta_0 = (ab)^{t^2-1}$, respectively [8, Lemma 2.1]. Therefore a (face-, vertex- or edge-) canonical metacyclic hypermap is chiral if and only if $t^2 \neq 1 \mod n$.

By *canonical metacyclic* we just mean face-canonical metacyclic.

In contrast, most of the hypermaps appearing in the classification [3] are not canonical metacyclic. However, we will show that all regular oriented hypermaps with a prime number of hyperfaces have metacyclic automorphism groups, though not necessarily being (face-, vertex- or edge-) canonical metacyclic hypermaps.

1.2 Primer hypermaps

We can use the equivalence of Proposition 11 of [4] for our definition of primer hypermap as it gives a good general idea behind the concept. A *(face-) primer* hypermap is a regular oriented hypermap with no non-trivial regular proper quotients with the same number of hyperfaces.

Any regular hypermap $\mathcal{H} = (G; a, b)$ covers a unique primer hypermap $\mathcal{P} = \mathcal{P}(\mathcal{H})$. In particular the hyperface valency l of its primer hypermap divides the hyperface valency n of $\mathcal{H}$. For consistency we reserve the letter ℓ to denote the valency of a hyperface of a primer hypermap and set aside the letter n for the valency of a hyperface of a non necessarily primer hypermap.

This primer hypermap can be constructed in the following way. When the elements of G act on G (set of darts) on the right they act as monodromy elements, but when they act on the left they act as automorphisms of $\mathcal{H}$. Therefore each element $\gamma \in G$ induces an automorphism $\varphi_\gamma : g \mapsto \gamma g$ of $\mathcal{H}$. In particular the automorphisms $\varphi_a : g \mapsto ag$ and $\varphi_b : g \mapsto bg$, induced by a and b, correspond to one-step global counter-clockwise rotations about the hyperface and the hypervertex (respectively) that contain the identity dart. Since $\mathcal{H}$ is regular we have $Aut(\mathcal{H}) = \langle \varphi_a, \varphi_b \rangle \cong G$ and since functions in this paper act on the right, $\varphi_{\gamma_1\gamma_2} = \varphi_{\gamma_2}\varphi_{\gamma_1}$ and thus $\varphi_{(\gamma_1\gamma_2)^{-1}} = \varphi_{\gamma_1^{-1}}\varphi_{\gamma_2^{-1}}$, and so,

$$\mathcal{H} = (G; a, b) \cong (Aut(\mathcal{H}); (\varphi_a)^{-1}, (\varphi_b)^{-1})\,.$$

The action of $Aut(\mathcal{H})$ on $\mathcal{H}$ induces a transitive action of $Aut(\mathcal{H})$ on the set of the hyperfaces $\mathcal{F} = G/_l\langle a \rangle$ of $\mathcal{H}$, where the symbol $G/_lK$ represents the left cosets of K in G. Under this action, each $\varphi_\gamma \in Aut(\mathcal{H})$, or equivalently each $\gamma \in G$, determines a permutation $\pi_\gamma \in Sym(\mathcal{F})$ defined by $g\langle a \rangle \mapsto \gamma g \langle a \rangle$. In particular, the automorphisms φ_a and φ_b give rise to permutations $A = {\pi_a}^{-1}$ and $B = {\pi_b}^{-1}$ on $\mathcal{F}$. Labelling the hyperfaces of $\mathcal{H}$ by 1, 2, ..., F, the permutations A and B are elements of the symmetric group S_F. Let P be the subgroup of S_F generated by A and B. Then $\mathcal{P} = \mathcal{P}(\mathcal{H}) = (P; A, B)$ is the *primer hypermap* determined by $\mathcal{H}$. The subgroup P of S_F generated by A and B is called the (face-) *primer group* of $\mathcal{H}$. We note that we are not adopting the notation $\mathcal{P} = (P; A^{-1}, B^{-1})$ we have used in [4].

The function $\Pi : G \longrightarrow P$, $\gamma \mapsto \gamma\Pi = {\pi_\gamma}^{-1} = \pi_{\gamma^{-1}}$ which maps $a \mapsto A$ and $b \mapsto B$, is an epimorphism with kernel $Kern(\Pi) = \langle a^{|A|} \rangle$ (Proposition 5 of [4]). Therefore it induces an epimorphism $\Pi : \mathcal{H} \longrightarrow \mathcal{P}$ branched over hyperfaces. Moreover, since $P \cong G/\langle a^{|A|} \rangle$, we get:

Corollary 7 [4]: *for any word* $r(A, B)$ *on* A, B, $r(A, B) = 1$ *if and only if* $r(a, b) = a^u$ *for some* $u = 0 \mod |A|$".

To recognise a canonical metacyclic hypermap from its primer we have the following proposition:

Proposition 4 [4]: $\mathcal{H}$ *is (face-) canonical metacyclic if and only if* $A = 1$.

The chirality group of the primer hypermap is a factor group of the chirality group of the hypermap (Proposition 9 of [4]), that is, $X(\mathcal{P}(\mathcal{H})) = X(\mathcal{H})/K$ for some K. Consequently, the chirality index $\kappa(\mathcal{P}(\mathcal{H}))$ divides $\kappa(\mathcal{H})$. Hence if $\mathcal{P}(\mathcal{H})$ is chiral then also $\mathcal{H}$ is chiral. The converse is not true.

The main theorem of [4] says:

Theorem 16 [4]: $\mathcal{P}$ *is a primer hypermap with* p *(prime) hyperfaces (each of valency* ℓ*) if and only if* $\mathcal{P} \cong \mathcal{P}_k^{p,\ell,t} = (M(p, \ell, 0, t); y, xy^k)$ *for some* $\ell, t \in \{1, \ldots, p-1\}$ *and* $k \in \{0, ..., \ell - 1\}$ *such that: (1)* ℓ *is a divisor of* $p - 1$, *(2)* $t^\ell = 1 \mod p$, *(3) if* $\ell > 1$, $t^i \neq 1 \mod p$ *for each* $i \in \{1, 2, ..., \ell - 1\}$. *Here* $M(p, \ell, 0, t)$ *is the metacyclic group* $\langle x, y \mid x^p = y^\ell = 1, x^y = x^t \rangle = \langle x \rangle \rtimes \langle y \rangle$. *Different parameters* k, ℓ *and* t *correspond to non-isomorphic primer hypermaps with* p *hyperfaces of valency* ℓ.

Note that if $\mathcal{H}$ is a regular oriented hypermap with p (not necessarily prime) hyperfaces, each of valency n, then $\mathcal{H}$ covers a unique primer hypermap $\mathcal{P}$ with p hyperfaces, each of valency ℓ with ℓ dividing n. Also useful is the following corolloary:
Corollary 17 [4]: *The H-sequences of the primer hypermaps $\mathcal{P}_k^{p,\ell,t}$ given above are*

(1) $[p, p, 1\,;\, 1, 1, p\,;\, p]$ *if* $k = 0$ *and* $\ell = 1$ $(\Rightarrow t = 1)$*;*

(2) $[p, \ell, \ell\,;\, \ell, p, p\,;\, \ell p]$ *if* $k = 0$ *and* $\ell > 1$ $(\Rightarrow p > 2)$*;*

(3) $[\ell, p, \ell\,;\, p, \ell, p\,;\, \ell p]$ *if* $k = \ell - 1 > 0$ $(\Rightarrow p > 2)$*;*

(4) $\left[\frac{\ell}{(\ell, k)}, \frac{\ell}{(\ell, k+1)}, \ell\,;\, p(\ell, k), p(\ell, k+1), p\,;\, \ell p\right]$ *if* $0 < k < \ell - 1$ $(\Rightarrow p > 2)$*,*

where, for space saving, (u, v) stands for $\gcd(u, v)$, the greatest common divider of u and v.

2 Derivations

Before we start with the classification, we introduce several families of regular oriented hypermaps that are derived from a given hypermap. These families together with their properties will be useful later on.

Let $\mathcal{H} = (G; a, b)$ be a regular (oriented) hypermap with F hyperfaces of valency n. The following regular hypermaps, which we call *derivations* of $\mathcal{H}$, all have the same number of hyperfaces F, and the same hyperface-valency n.

(1) The *mirror* $\overline{\mathcal{H}} = (G; a^{-1}, b^{-1})$;

(2) The *mid-mirror* $Mm(\mathcal{H}) := (G; a, b^{-1})$;

(3) The *k-Left* family $L_k(\mathcal{H}) := (G; a, a^k b)$, for each $k \in \{1, ..., n-1\}$,

(4) The *k-Right* family $R_k(\mathcal{H}) := L_k(\mathcal{H})^{a^k} = (G; a^{a^k}, (a^k b)^{a^k}) = (G; a, ba^k)$, and

(5) The $(0,1)$-*dual* $D_{(0,1)}(\mathcal{H}) = Mm(L_1(\mathcal{H})) = (G; a, (ab)^{-1})$; this is the hypermap resulting from $\mathcal{H}$ by swapping hypervertices with hyperedges.

One easily sees that $\overline{\overline{\mathcal{H}}} = \mathcal{H}$, $Mm(Mm(\mathcal{H})) = \mathcal{H}, D_{(0,1)}(D_{(0,1)}(\mathcal{H})) = \mathcal{H}$, $L_{n-k}(L_k(\mathcal{H})) = L_k(L_{n-k}(\mathcal{H})) = \mathcal{H}$ and $R_{n-k}(R_k(\mathcal{H})) = R_k(R_{n-k}(\mathcal{H})) = \mathcal{H}$.

Let $\mathcal{D}(\mathcal{H})$ denote one of the derivations of $\mathcal{H}$. Then $\mathcal{D}$ defines an operation $\mathcal{D} : \mathcal{H} \mapsto \mathcal{D}(\mathcal{H})$ that takes a regular oriented hypermap with F hyperfaces of valency n to a regular oriented hypermap with F hyperfaces of valency n. This operation has the inverse defined by

$$\mathcal{D}^{-1} = \begin{cases} \mathcal{D} & \text{if } \mathcal{D} \text{ is the mirror, mid-mirror or the } (0,1)\text{-dual} \\ L_{n-k} & \text{if } \mathcal{D} = L_k \\ R_{n-k} & \text{if } \mathcal{D} = R_k\,. \end{cases}$$

Denote by Π, $\overline{\Pi}$, Π_M, Π_L and Π_R the corresponding homomorphisms $G \longrightarrow S_p$, $\gamma \mapsto \pi_{\gamma^{-1}}$. For example, $\Pi : a \mapsto \pi_{a^{-1}}$, $b \mapsto \pi_{b^{-1}}$ and $\Pi_L : a \mapsto \pi_{a^{-1}}$, $a^k b \mapsto \pi_{b^{-1}a^{-k}}$. As $b\Pi_L = (a^{-k} a^k b)\Pi_L = a^{-k}\Pi_L\, a^k b\Pi_L = \pi_{a^k}\pi_{b^{-1}a^{-k}} = \pi_{b^{-1}a^{-k}\,a^k} = \pi_{b^{-1}}$, then $\Pi_L = \Pi$. Similarly we have $\Pi = \overline{\Pi} = \Pi_L = \Pi_M$. Since the primer hypermap of $\mathcal{H}$ is $\mathcal{P}(\mathcal{H}) = (G\Pi; a\Pi, b\Pi)$ and $\mathcal{H}$ is primer if and only if $Ker(\Pi) = 1$, we immediately have,

Proposition 2.1. *Let $\mathcal{H}$ be a regular hypermap and $\mathcal{P}(\mathcal{H})$ be its primer hypermap. Then*

(1) $\mathcal{P}(\mathcal{D}(\mathcal{H})) = \mathcal{D}(\mathcal{P}(\mathcal{H}))$, *for any derivation* $\mathcal{D}$ *of* $\mathcal{H}$.

(2) $\mathcal{H}$ *is primer if and only if any of its derivations is primer*.

Let $\mathcal{F}_\mathcal{P}$ denote the family of regular hypermaps with primer hypermap $\mathcal{P}$. As an immediate consequence of above, $\mathcal{D}(\mathcal{F}_\mathcal{P}) = \mathcal{F}_{\mathcal{D}(\mathcal{P})}$, and, as a consequence of this, we have $\mathcal{D}(\mathcal{H}) \in \mathcal{F}_\mathcal{P} \Leftrightarrow \mathcal{H} \in \mathcal{F}_{\mathcal{D}^{-1}(\mathcal{P})}$.

Let $\mathcal{H} = (G; a, b)$ be a regular hypermap and $\langle a, b \mid R(a, b)\rangle$ be a presentation for the monodromy group G. Let $x = X(a, b)$ and $y = Y(a, b)$ be another pair of generators of G. Then the original generators a and b can be written as words in x and y, say $a = A(x, y)$ and $b = B(x, y)$.

Proposition 2.2. *If the change of generators a to x and b to y produce no extra relations, that is, if $x = X(A(x, y), B(x, y))$ and $y = Y(A(x, y), B(x, y))$ are not new relations, and there is a $w \in G$ such that the conjugations $A^w = w^{-1}Aw$ and $B^w = w^{-1}Bw$ coincide with their inverse order words, in symbols $A^w = \overleftarrow{A}$ and $B^w = \overleftarrow{B}$, then both hypermaps $\mathcal{H} = (G; a, b)$ and $\mathcal{Q} = (G; x, y)$ have the same chirality group.*

Proof. The non-existence of any extra relations implies that $\langle x, y \mid R\big(A(x, y), B(x, y)\big)\rangle$ is another presentation of G, this time as a function of the new generators x and y. By Theorem 1 of [2] we have,

$$\begin{aligned} X(\mathcal{Q}) &= \langle R(A(x^{-1}, y^{-1}), B(x^{-1}, y^{-1}))\rangle^G = \langle R(\overleftarrow{A}(x, y)^{-1}, \overleftarrow{B}(x, y)^{-1})\rangle^G \\ &= \langle R(A^w(x, y)^{-1}, B^w(x, y)^{-1})\rangle^G \\ &= \langle R(A(x, y)^{-1}, B(x, y)^{-1})\rangle^G = \langle R(a^{-1}, b^{-1})\rangle^G \\ &= X(\mathcal{H}). \end{aligned}$$

□

We saw in Proposition 2.1 that $\mathcal{H}$ is primer if and only if any of its derivations $\mathcal{D}(\mathcal{H})$ is also primer. Now we show that this is also true for chirality.

Corollary 2.3. *Let $\mathcal{D}$ be a derivation of $\mathcal{H}$. Then $X(\mathcal{D}(\mathcal{H})) = X(\mathcal{H})$; that is $\mathcal{H}$ and its derivations $D(\mathcal{H})$ all share the same chirality group. In particular, if some derivation of $\mathcal{H}$ is chiral then also $\mathcal{H}$ is chiral.*

Proof. In Proposition 2.2 take $w = id$ if $\mathcal{Q} = \overline{\mathcal{H}}$ or $Mm(\mathcal{H})$, and take $w = a^{-k}$ if $\mathcal{Q} = L_k(\mathcal{H})$ or $R_k(\mathcal{H})$. □

Consider the families $\mathcal{P}^p_I = \{\mathcal{P}^{p,1,1}_0\}$, $\mathcal{P}^p_{II} = \{\mathcal{P}^{p,\ell,t}_0\}_{\ell,t}$ with $\ell > 1$, $\mathcal{P}^p_{III} = \{\mathcal{P}^{p,\ell,t}_{\ell-1}\}_{\ell,t}$ with $\ell > 1$, and $\mathcal{P}^p_{IV} = \{\mathcal{P}^{p,\ell,t}_k\}_{k,\ell,t}$ with $\ell > 1$ and $0 < k < \ell-1$, of the primer hypermaps with H-sequences (1), (2), (3) and (4) respectively, of Corollary 17 [4]. Then $\mathcal{P}^p_{III} = R_{\ell-1}(\mathcal{P}^p_{II})$ and $\mathcal{P}^p_{IV} = R_k(\mathcal{P}^p_{II})$. Taking into account Corollary 2.3 and [7, Corollary 9] for the chirality group of canonical metacyclic hypermap, we have the following result shown in [4]:

Corollary 2.4. *If $\mathcal{P} = \mathcal{P}^{p,\ell,t}_k = (G; y, xy^k)$ is a primer hypermap with p hyperfaces (p prime) then*

$$X(\mathcal{P}) = \begin{cases} 1 \text{ (reflexible)} & \text{if } \mathcal{P} \in \mathcal{P}^p_I \\ \langle y^{t^2-1}\rangle & \text{if } \mathcal{P} \in \mathcal{P}^p_{II}, \mathcal{P}^p_{III} \text{ or } \mathcal{P}^p_{IV}\,. \end{cases}$$

3 The classification

Let p be a prime number. We now proceed with the classification of the regular oriented hypermaps with p hyperfaces. Let $\mathcal{H} = (G; a, b)$ be a regular oriented hypermap with p hyperfaces (of valency n) and $\mathcal{P} = (G\Pi; a\Pi, b\Pi) = (P; A, B)$ be its primer hypermap, which also has p hyperfaces. In what follows,

$$M(n, p, u, t) := \langle a, b \mid a^n = 1,\, b^p = a^u,\, b^{-1}ab = a^t \rangle$$

is the metacyclic group with parameters n, p, u, t, and

$$G^{p,\ell,t}_{n,u,v} := \langle a, b \mid a^n = 1,\, b^p = a^u,\, [a^\ell, b] = 1, bab^{-t} = a^v \rangle\,.$$

Before we state and prove the main theorem, we first prove the following lemma,

Lemma 3.1. *Let $G = \langle a, b\rangle$, p an odd prime and t, ℓ positive integers such that $t \neq 1 \mod p$, $t^\ell = 1 \mod p$ and $p = 1 \mod \ell$. If (i) $b^p \in \langle a^\ell\rangle$, (ii) $b^i ab^{-it} \in \langle a^\ell\rangle a$, for any $i = 1, 2, \ldots, p$, then $b \leftrightharpoons a^\ell$, where the symbol $\leftrightharpoons$ means "commutes with".*

Proof. As $t^\ell = 1 \mod p$, by (i) we have $b^{t^\ell - 1} \in \langle a^\ell\rangle$. On the other hand, taking $i = 1, t, t^2, \ldots, t^{\ell-1}$, (ii) yields the following information bab^{-t}, $b^t ab^{-t^2}$, ..., $b^{t^{\ell-1}} ab^{-t^\ell} \in \langle a^\ell\rangle a$. Multiplying the first i of these words (in the same order as shown) and the last $\ell - i$ words we get $ba^i b^{-t^i} \in \langle a^\ell\rangle a^i$ and $b^{t^i} a^{\ell-i} b^{-t^\ell} \in \langle a^\ell\rangle a^{\ell-i}$, respectively. Then $ba^\ell b^{-t^\ell} \in \langle a^\ell\rangle$ and thus

$$ba^\ell b^{-1} = (ba^\ell b^{-t^\ell})\, b^{t^\ell - 1} = a^V$$

for some $V = 0 \mod \ell$. On the other hand,

$$b^{t^i} a^\ell b^{-t^i} = (b^{t^i} a^{\ell-i} b^{-1})(ba^i b^{-t^i}) = (ba^i b^{-t^i})(b^{t^i} a^{\ell-i} b^{-1}) = ba^\ell b^{-1} = a^V$$

for each $i \in \{0, \ldots, \ell-1, \ell\}$. Consequently, $b^{(t^i - 1)} a^\ell b^{1 - t^i} = b^{-1} a^V b = a^\ell$ for every integer $i \in \{0, \ldots, \ell - 1, \ell\}$. Taking $i = \ell$ we get

$$[a^\ell, b^{t^i - 1}] = 1, \;\; i = 0, \ldots, \ell.$$

In particular,

$$b^{(t-1)+(t^2-1)+\ldots+(t^{\ell-1}-1)} = b^{(1+t+t^2+\ldots+t^{\ell-1})-\ell} \leftrightharpoons a^\ell\,. \tag{3.1}$$

As $t \neq 1 \mod p$ and $t^\ell - 1 = (t-1)(1 + t + t^2 + \ldots + t^{\ell-1}) = 0 \mod p$, one has $(1 + t + t^2 + \ldots + t^{\ell-1}) = 0 \mod p$, and so by (3.1) we have $b^\ell \leftrightharpoons a^\ell$. As ℓ is a divisor of $p - 1$, we also have $b^{p-1} \leftrightharpoons a^\ell$. Consequently

$$[a^\ell, b] = 1.$$

□

Theorem 3.2. *If $\mathcal{H} = (G; a, b)$ is a regular oriented hypermap with p (prime) hyperfaces, each of valency n, then $\mathcal{H}$ is isomorphic to one of the following hypermaps:*

(1) $\mathcal{CM}_{n,p,u,t} = (M(n,p,u,t);\, a\,,\, b)$ *for some* $u,\, t \in \{0,1,\ldots,n-1\}$ *such that*

$$(t-1)u = 0 \mod n \quad \textit{and} \quad t^p = 1 \mod n\,;$$

(2) $\mathcal{H}^{p,\ell,t,k}_{n,u,v} = (G^{p,\ell,t}_{n,u,v};\, a\,,\, ba^k)$ *(p odd prime) ,* *for some* $\ell \in \{2,\ldots,n\}$, $u, v \in \{0,\ldots,n-1\}$, $k \in \{0,\, ...,\, \ell-1\}$ *and* $t \in \{2,\ldots,p-1\}$ *such that*

- *(H1)* $gcd(p-1,n) = 0 \mod \ell$,
- *(H2)* $t^\ell = 1 \mod p$ *and* $t^i \neq 1 \mod p$ *for* $i \in \{1, 2,\, ...,\, \ell-1\}$ (*that is, t has order* ℓ *in* $\mathbb{Z}^*_p = \mathbb{Z}_p \backslash \{0\}$),
- *(H3)* $u = 0 \mod \ell$, $v = 1 \mod \ell$ *and*
- *(H4)* $(t-1)u + p(v-1) = 0 \mod n$.

Moreover, all these hypermaps $\mathcal{H}^{p,\ell,t,k}_{n,u,v}$ *for* ℓ, t, k, n, u, v *satisfying the above conditions, have p hyperfaces of valency n, and different parameters* (ℓ, t, k, u, v) *correspond to non-isomorphic hypermaps with p hyperfaces of valency n.*

Proof. Let $\mathcal{H} = (G; a, b)$ be a regular oriented hypermap with p (prime) hyperfaces and $\mathcal{P} = (P; A, B)$ be its primer hypermap where $A = a\Pi$ and $B = b\Pi$. By Theorem 16 of [4], $\mathcal{P} = \mathcal{P}^{p,\ell,t}_k$, for some k, ℓ and t. We treat separately the following cases: A=1 (Case 1), $A \neq 1$ and $|B| = p$ (Case 2) and $A \neq 1$ and $|B| \neq p$ (Case 3).

Case 1. If $A = 1$ then $\mathcal{H}$ is canonical metacyclic and $\mathcal{P} = \overline{\mathcal{P}}$ is the spherical cyclic hypermap $\mathcal{C}_p = (C_p; 1, B)$ (Proposition 4 of [4]). Then $\mathcal{H}$ is isomorphic to $\mathcal{CM}_{n,p,u,t} = (M(n,p,u,t); a, b)$ for some u and t such that $(t-1)u = 0 \mod n$ and $t^p = 1 \mod n$. ♦

Case 2. $|A| = \ell > 1$ and $|B| = p$. By Theorem 16 of [4], and Corollary 17 of [4] (see also §1.2), $\mathcal{P}(\mathcal{H}) = \mathcal{P}^{p,\ell,t}_0 = (P; A, B)$, where

$$P = \langle A, B \mid A^\ell = 1,\, B^p = 1,\, A^{-1}BA = B^t \rangle\,, \tag{3.2}$$

for some $t \in \{1,\ldots,p-1\}$ such that $t^\ell = 1 \mod p$ and $t^i \neq 1 \mod p$ for $i = 1, 2,\, ...,\, \ell-1$. By Proposition 15 of [4], ℓ is a divisor of $p-1$ (that is, $p = 1 \mod \ell$).

From $A^{-1}BA = B^t$ we deduce that $B^iAB^{-it} = A$ for each integer i. Applying Corollary 7 of [4] (see also §1.2), we derive the following relations in G:

- (i) $a^n = 1$ with $n = 0 \mod \ell$;
- (ii) $b^p = a^u$, $u = 0 \mod \ell$;
- (iii) $b^iab^{-it} = a^{v_i}$, $v_i = 1 \mod \ell$, $i = 1,\, ...,\, p-1$ (also true for $i = p$).

These equations define a group that is right factorised by $K = \langle a \rangle$ into p cosets. Indeed, on the one hand $K, Kb, \ldots, Kb^{p-1}$ are distinct cosets because p is the smallest positive integer such that b^p belongs to $K = \langle A \rangle$, and on the other hand, since $\{0, t,\, 2t,\, ...,\, (p-1)t\}$ is a complete set of residues modulo p, equation (iii) implies that the set of right cosets of K in G is $G/_rK = \{K,\, Kb,\, ...,\, Kb^{p-1}\}$. Hence the monodromy group of $\mathcal{H}$ is given by

$$G = \langle a, b \mid a^n = 1,\, b^p = a^u,\, b^iab^{-it} = a^{v_i},\, i = 1,\, ...,\, p-1 \rangle$$

for some integers $n = 0 \mod \ell$, $u = 0 \mod \ell$ and $v_i = 1 \mod \ell$, $i = 1, ..., p-1$. We now simplify this presentation.

From relation (iii) we have $b^i ab^{-it} \in \langle a^\ell \rangle a$, valid for any i. By Lemma 3.1 ,

$$[a^\ell, b] = 1.$$

Adding the relation $[a^\ell, b] = 1$ to the presentation of G, some relations of the previous presentation will turn out to be redundant. From (iii), $i = 1$, we deduce $b^2ab^{-2t} = b(bab^{-t})b^{-t} = ba^{v_1}b^{-t} = a^{v_1-1}bab^{-t} = a^{2v_1-1} = a^{2(v_1-1)+1}$; and more generally

$$b^i ab^{-it} = a^{i(v_1-1)+1},\ i = 1, ..., p-1 \text{ and } p.$$

Thus all the relations in (iii) except the first one are redundant. Now for $i = p$ we also have $b^p ab^{-pt} = a^{(1-t)u+1}$; therefore $a^{(1-t)u+1} = a^{p(v_1-1)+1}$ which implies

$$(1-t)u = p(v_1 - 1) \mod n.$$

The hypermap $\mathcal{H}$ is then isomorphic to $\mathcal{H}^{p,\ell,t,0}_{n,u,v} := (G^{p,\ell,t}_{n,u,v}; a, b)$, where

$$G^{p,\ell,t}_{n,u,v} = \langle a, b \mid a^n = 1,\ b^p = a^u,\ [a^\ell, b] = 1, bab^{-t} = a^v \rangle\,,$$

for some ℓ, t, n, u and v such that $p = 1 \mod \ell$, $t^\ell = 1 \mod p$, $t^i \neq 1 \mod p$ for each $i \in \{1, 2, ..., \ell-1\}$, $u = 0 \mod \ell$, $v = 1 \mod \ell$, $n = 0 \mod \ell$ and $(1-t)u = p(v-1) \mod n$.

Conversely, we show that if $\mathcal{H} = \mathcal{H}^{p,\ell,t,0}_{n,u,v}$ for some ℓ, t, n, u, v satisfying the above conditions, then $\mathcal{H}$ has p hyperfaces of valency n. Factoring $G = G^{p,\ell,t}_{n,u,v}$ by the normal subgroup $\langle a^\ell \rangle$ yields the monodromy group of the primer hypermap $\mathcal{P} = \mathcal{P}^{p,\ell,t}_0$ with p hyperfaces of valency ℓ. Then ℓ divides $|a|$ and so both $\mathcal{P} = \mathcal{H}/\langle a^\ell \rangle$ and $\mathcal{H}$ have the same number of hyperfaces, p.

As $\gcd(t-1, p) = 1$ there exist integers c and d such that $c(t-1) + dp = 1$. Then the assignment $a \mapsto a$, $b \mapsto a^{c(1-v)+du}$ turn each of the relators of G into the identity of $C_n = \langle a \mid a^n = 1 \rangle$ and so it defines an epimorphism from G to the cyclic group C_n. This proves that the order of a in G is n. Consequently $\mathcal{H}$ has p hyperfaces of valency n. ♦

Case 3. $|A| = \ell > 1$ and $|B| \neq p$. By Theorem 16 of [4], and Corollary 17 of [4],

$$\mathcal{P}(\mathcal{H}) = \mathcal{P}^{p,\ell,t}_k = (M(p,\ell,0,t); A, \beta A^k) = R_k((M(p,\ell,0,t); A, \beta)) = R_k(\mathcal{P}^{p,\ell,t}_0)$$

for some $k \in \{0, ..., \ell-1\}$ and $t \in \{1, 2, \ldots, p-1\}$, where $M(p,\ell,0,t) = \langle \beta, A \mid \beta^p = A^\ell = 1, \beta^A = \beta^t \rangle = \langle \beta \rangle \rtimes \langle A \rangle$ and (1) ℓ is a divisor of $p-1$, (2) $t^\ell = 1 \mod p$ and (3) if $\ell > 1$, $t^i \neq 1 \mod p$ for each $i \in \{1, 2, ..., \ell-1\}$.

Then $\mathcal{P}(R_{\ell-k}(\mathcal{H})) = \mathcal{P}^{p,\ell,t}_0$. By case 2, $R_{\ell-k}(\mathcal{H}) = \mathcal{H}^{p,\ell,t,0}_{n,u,v} = (G^{p,\ell,t}_{n,u,v}; a, b)$. Thus

$$\mathcal{H} = \mathcal{H}^{p,\ell,t,k}_{n,u,v} := R_k(\mathcal{H}^{p,\ell,t,0}_{n,u,v}) = (G^{p,\ell,t}_{n,u,v};\ a\,,\ ba^k)\,,$$

for some $k \in \{0, ..., \ell-1\}$. ♦

Finally we show that different parameters lead to different hypermaps.

(1) If $\mathcal{CM}_{n,p,u,t} = (M; a, b) \cong \mathcal{CM}_{n',p',u',t'} = (M'; \alpha, \beta)$ then $p' = p$ and $n' = n$, since the number of hyperfaces and their valencies should be the same. Then it becomes obvious that we must have $u' = u$ and $t' = t$ since these parameters run from 0 to $n-1$.
(2) If $\mathcal{H}^{p,\ell,t,k}_{n,u,v} \cong \mathcal{H}^{p',\ell',t',k'}_{n',u',v'}$, then, as above, we must have $p' = p$ and $n' = n$. Looking at the primer hypermaps,

$$\mathcal{P}(\mathcal{H}^{p,\ell,t,k}_{n,u,v}) = \mathcal{P}^{p,\ell,t}_{k} \cong \mathcal{P}(\mathcal{H}^{p,\ell',t',k'}_{n,u',v'}) = \mathcal{P}^{p,\ell,t'}_{k'},$$

this forces $\ell' = \ell$, $t' = t$ and $k' = k$ (Theorem 16 [4]). But then we have $\mathcal{H}^{p,\ell,t,k}_{n,u,v} = R_k(\mathcal{H}^{p,\ell,t,0}_{n,u,v})$ and $\mathcal{H}^{p,\ell,t,k}_{n,u',v'} = R_k(\mathcal{H}^{p,\ell,t,0}_{n,u',v'})$, so we must also have $\mathcal{H}^{p,\ell,t,0}_{n,u,v} \cong \mathcal{H}^{p,\ell,t,0}_{n,u',v'}$. It is now clear that this isomorphism implies $u' = u$ and $v' = v$, since u, u', v, v' are restrict to $\{0, \ldots, n-1\}$. □

Each hypermap $\mathcal{H}$ with p hyperfaces, p prime, with each hyperface of valency n, covers only one primer hypermap also with p hyperfaces of valency $\ell = |A|$ (a divisor of n). This primer hypermap is

$$\mathcal{P}(\mathcal{CM}_{n,p,u,t}) = \mathcal{C}_p = (C_p; 1, B), \text{ if } \mathcal{H} = \mathcal{CM}_{n,p,u,t}, \text{ or}$$
$$\mathcal{P}(\mathcal{H}^{p,\ell,t,k}_{n,u,v}) = \mathcal{P}^{p,\ell,t}_{k} = (G^{p,\ell,t}_{0}; a, ba^k), \text{ if } \mathcal{H} = \mathcal{H}^{p,\ell,t,k}_{n,u,v}.$$

Among the hypermaps in the family $\mathcal{H}^{p,\ell,t,k}_{n,u,v}$ many of them share the same group and are distinguished by different pairs of generators. In the previous proof we find the requisites necessary to prove that $G^{p,\ell,t}_{n,u,v}$ is a metacyclic group.

Proposition 3.3. *$G^{p,\ell,t}_{n,u,v}$ is a metacyclic group isomorphic to $G^{p,\ell,t}_{n,0,1} = M(p,n,0,t) = \langle \beta, \alpha \mid \beta^p = 1,\ \alpha^n = 1,\ \alpha^{-1}\beta\alpha = \beta^t \rangle$ under the isomorphism $\psi : a \mapsto \alpha$, $b \mapsto \beta\alpha^\theta$, where $\theta = c(1-v) + du$, for some c, d satisfying $c(t-1) + dp = 1 = \gcd(t-1, p)$. Moreover, $\mathcal{H}^{p,\ell,t,k}_{n,u,v} \cong R_{\theta+k}(\mathcal{H}^{p,\ell,t}_{n})$, where $\mathcal{H}^{p,\ell,t}_{n}$ is the canonical metacyclic hypermap $(G^{p,\ell,t}_{n,0,1}; \alpha, \beta)$.*

Proof. Consider the group

$$\begin{aligned} G^{p,\ell,t}_{n,0,1} &= \langle \alpha, \beta \mid \alpha^n = 1,\ \beta^p = 1,\ \beta\alpha\beta^{-t} = \alpha \rangle \\ &= \langle \beta, \alpha \mid \beta^p = 1,\ \alpha^n = 1,\ \alpha^{-1}\beta\alpha = \beta^t \rangle \\ &= M(p, n, 0, t). \end{aligned}$$

Note that $\beta\alpha\beta^{-t} = \alpha \Leftrightarrow \alpha^{-1}\beta\alpha = \beta^t$ implies that $\beta^i\alpha\beta^{-it} = \alpha$ for any i; so by Lemma 3.1, $[\alpha^\ell, \beta] = 1$. This group, being metacyclic, has order np. Note also that the condition $t^\ell = 1 \mod p$ is stronger than the metacyclic condition $t^n = 1 \mod p$. Let c and d be such that $c(t-1) + dp = 1$, and let $\theta = c(1-v) + du$. Then

$$\theta = 0 \mod \ell, \text{ so } \alpha^\theta \in Z(G^{p,\ell,t}_{n,0,1}),$$

$$\theta p = cp(1-v) + pdu = c(t-1)u + pdu = (c(t-1) + pd)u = u, \quad \text{and}$$

$$\begin{aligned} \theta(1-t) &= (1-t)c(1-v) + (1-t)u\,d \\ &= (1-t)c(1-v) + p(v-1)d \\ &= (t-1)c(v-1) + p(v-1)d \\ &= (v-1)((t-1)c + pd) \\ &= v-1. \end{aligned}$$

The assignment $\psi : a \mapsto \alpha$, $b \mapsto \beta\alpha^{\theta}$, transfers the relators of $G^{p,\ell,t}_{n,u,v}$ to relators of $G^{p,\ell,t}_{n,0,1}$ as we can observe:

$$\begin{aligned}
&1) \quad a^n \longrightarrow \alpha^n = 1,\\
&2) \quad b^p a^{-u} \longrightarrow (\beta\alpha^{\theta})^p \alpha^{-u} = \beta^p \alpha^{\theta p}\alpha^{-u} = \alpha^u\alpha^{-u} = 1,\\
&3) \quad [a^{\ell}, b] \longrightarrow [\alpha^{\ell}, \beta\alpha^{\theta}] = [\alpha^{\ell}, \beta] = 1,\\
&4) \quad bab^{-t}a^{-v} \longrightarrow \beta\alpha^{\theta}\alpha(\beta\alpha^{\theta})^{-t}\alpha^{-v} = \beta\alpha\beta^{-t}\alpha^{\theta}\alpha^{-\theta t}\alpha^{-v}\\
&\qquad = \alpha^{\theta(1-t)}\alpha\alpha^{-v} = \alpha^{v-1}\alpha^{1-v} = 1,
\end{aligned}$$

By the Substitution Test [15, Theorem 4, pg 29], $\psi : G^{p,\ell,t}_{n,u,v} \to G^{p,\ell,t}_{n,0,1}$ is an epimorphism. As $|G^{p,\ell,t}_{n,0,1}| = |G^{p,\ell,t}_{n,u,v}|$, ψ is indeed an isomorphism and thus $G^{p,\ell,t}_{n,u,v}$ is metacyclic. This isomorphism also shows that $\mathcal{H}^{p,\ell,t}_{n,u,v} = \mathcal{H}^{p,\ell,t,0}_{n,u,v}$ is isomorphic to $R_{\theta}(\mathcal{H}^{p,\ell,t}_{n})$. Then $\mathcal{H}^{p,\ell,t,k}_{n,u,v} = R_k(\mathcal{H}^{p,\ell,t}_{n,u,v}) \cong R_{\theta+k}(\mathcal{H}^{p,\ell,t}_{n})$. □

Corollary 3.4. *If $\mathcal{H}$ is a regular oriented hypermap with a prime number of hyperfaces then its automorphism group is metacyclic, though $\mathcal{H}$ is not necessarily canonical metacyclic.*

4 Chirality groups and H-sequences

Theorem 4.1. *The chirality groups of $\mathcal{CM}_{n,p,u,t}$ and $\mathcal{H}^{p,\ell,t,k}_{n,u,v}$ are the cyclic groups $\langle a^{t^2-1}\rangle$ and $\langle b^{t^2-1}\rangle$ respectively. The chirality index of $\mathcal{CM}_{n,p,u,t}$ is $\frac{n}{(n,t^2-1)}$ while the chirality index of $\mathcal{H}^{p,\ell,t,k}_{n,u,v}$ is*

$$\frac{p}{\gcd(p, t^2-1)} = \begin{cases} 1, & t = p-1\\ p, & t \in \{2, \dots, p-2\}\end{cases}$$

Proof. The canonical metacyclic hypermap $\mathcal{CM}_{n,p,u,t}$ has chirality group $\langle a^{t^2-1}\rangle$ [7].

By Proposition 3.3, $\mathcal{H}^{p,\ell,t,k}_{n,u,v} = R_{\theta+k}\left(\mathcal{H}^{p,\ell,t}_{n}\right)$ and, by Proposition 2.3, $\mathcal{H}^{p,\ell,t,k}_{n,u,v}$ has the same chirality group as the vertex-canonical metacyclic hypermap $\mathcal{H}^{p,\ell,t}_{n} = (G^{p,\ell,t}_{n}; a, b)$, where

$$G^{p,\ell,t}_{n} = M(p, n, 0, t) = \langle b, a \mid b^p = 1,\ a^n = 1,\ a^{-1}ba = b^t\rangle .$$

Hence $\mathcal{H}^{p,\ell,t,k}_{n,u,v}$ has chirality group $\langle b^{t^2-1}\rangle$ [7], a subgroup of the cyclic group $\langle b\rangle$ of order p (prime). If $\mathcal{H}^{p,\ell,t,k}_{n,u,v}$ is not reflexible ($b^{t^2-1} \neq 1 \Leftrightarrow t \neq -1 \mod p$, since $t \geq 2$), it must has order p, and thus $\mathcal{H}^{p,\ell,t,k}_{n,u,v}$ has chirality index $\kappa = p$. □

Corollary 4.2. *If $\mathcal{H}$ is a regular oriented hypermap with a prime number of hyperfaces and its primer hypermap is $\mathcal{P} = (P; A, B)$, with $A \neq 1$, then $\mathcal{H}$ is reflexible if and only if $|A| = 2$.*

Theorem 4.3. *The H-sequences of the hypermaps of Theorem 3.2 are*

$$\mathcal{CM}_{n,p,u,t} : \left[\frac{pn}{(n,u)}, \frac{pn}{(n,t^{p-1}+...+1+u)}, n\,; (n,u), (n,t^{p-1}+...+1+u), p\,; pn\right]$$

$$\mathcal{H}^{p,\ell,t,0}_{n,u,v} : \left[\frac{pn}{(n,u)}, \frac{n}{(n,\theta+1)}, n\,; (n,u), p(n,\theta+1), p\,; pn\right]$$

$$\mathcal{H}^{p,\ell,t,\ell-1}_{n,u,v} : \left[\frac{n}{(n,\theta+\ell-1)}, ((\frac{pn}{(n,u)}, \frac{n}{\ell})), n\,; p(n,\theta+\ell-1), \frac{pn}{((\frac{pn}{(n,u)}, \frac{n}{\ell}))}, p\,; pn\right]$$

$$\underset{(0<k<\ell-1)}{\mathcal{H}^{p,\ell,t,k}_{n,u,v}} : \left[\frac{n}{(n,\theta+k)}, \frac{n}{(n,\theta+k+1)}, n\,; p(n,\theta+k), p(n,\theta+k+1), p\,; pn\right],$$

where, for space saving, (u,v) *stands for* $\gcd(u,v)$*,* $((u,v))$ *stands for* $\operatorname{lcm}(u,v)$*, the least common multiple of* u *and* v*,* $\theta = c(1-v)+du$*, and* c *and* d *are integers such that* $c(t-1)+dp = 1 = (t-1,p)$*.*

Proof. Recall that the H-sequence of $\mathcal{H} = (G;a,b)$ is a sequence of numbers

$$[|b|, |ab|, |a|; V, E, F; |G|]$$

where $|b|$ (the order of b) is the hypervertex-valency, $|ab|$ is the hyperedge-valency, $|a|$ is the hyperface-valency, $V = \frac{|G|}{|b|}$ is the number of hypervertices, $E = \frac{|G|}{|ab|}$ is the number of hyperedges and $F = \frac{|G|}{|a|}$ is the number of hyperfaces.

- For the canonical metacyclic hypermap $\mathcal{CM}_{n,p,u,t} = (G;a,b)$, where

$$G = M(n,p,u,t) = \langle a, b \mid a^n = 1,\, b^p = a^u,\, b^{-1}ab = a^t\rangle,$$

  it is clear that $|a| = n$, $|b| = \frac{pn}{\gcd(n,u)}$ and $|G| = pn$. We only need to calculate $|ab| = |ba|$. From the relation $b^{-1}ab = a^t$ we get: $a^m b = ba^{mt}$ and $ab^m = b^m a^{t^m}$, for any positive integer m. From this we easily derive

$$(ba)^m = b^m a^{t^{m-1}+\cdots+t+1}.$$

  Now as $\mathcal{CM}_{n,p,u,t}$ covers $\mathcal{P} = (C_p;1,B)$, the order of ba is a multiple of p. As $(ba)^p = a^{t^{p-1}+\cdots+t+1+u}$, then $|ba| = \frac{pn}{\gcd(n,t^{p-1}+...+1+u)}$. The rest of the H-sequence is easily determined from these values.
- For $\mathcal{H} = \mathcal{H}^{p,\ell,t,0}_{n,u,v} = (G;a,b)$, where $G = G^{p,\ell,t}_{n,u,v}$, we also have $|a| = n$, $|b| = \frac{pn}{\gcd(n,u)}$ and by Proposition 3.3, $|G| = pn$. Therefore we only need to calculate $|ab| = |ba|$.
- For $\mathcal{H} = \mathcal{H}^{p,\ell,t,\ell-1}_{n,u,v} = (G;a,ba^{\ell-1})$, where $G = G^{p,\ell,t}_{n,u,v}$, we have $|a| = n$, $|aba^{\ell-1}| = |ba^{\ell}| = \operatorname{lcm}(|b|,|a^{\ell}|) = \operatorname{lcm}(\frac{pn}{\gcd(n,u)}, \frac{n}{\ell})$, where $|G| = pn$. Therefore we only need to calculate $|ba^{\ell-1}|$.
- For $\mathcal{H} = \mathcal{H}^{p,\ell,t,k}_{n,u,v} = (G;a,ba^k)$, where $G = G^{p,\ell,t}_{n,u,v}$, we have $|a| = n$, $|aba^k| = |ba^{k+1}|$ and $|G| = pn$. Therefore we only need to calculate $|ba^{k+1}|$.

To complete the H-sequence in the last three cases, we need to calculate the order of $a^j b$ for $j \in \{1, ..., \ell-1\}$, within the group $G = G^{p,\ell,t}_{n,u,v}$. This group is isomorphic to the metacyclic group $M(p,n,0,t) = \langle \beta, \alpha \mid \beta^p = 1, \alpha^n = 1, \beta^\alpha = \beta^t \rangle$ under the isomorphism $\psi : G \to M(p,n,0,t)$, $a \mapsto \alpha$, $b \mapsto \beta\alpha^\theta$ (Proposition 3.3), where $\theta = c(1-v)+du$ (which is a multiple of ℓ), and c and d are integers such that $c(t-1)+dp = 1$. Then the order of $a^j b$ is the order of $\alpha^j \beta \alpha^\theta = \alpha^{\theta+j}\beta = \alpha^i \beta$, where $i = \theta + j \neq 0 \mod \ell$. We now follow the proof of Corollary 17 of [4] to compute the order of $\beta\alpha^i$.

The third equation of $M(p,n,0,t)$ implies that $\beta^{\alpha^i} = \beta^{t^i} \Leftrightarrow \beta\alpha^i = \alpha^i \beta^{t^i}$. By induction we get

$$(\beta\alpha^i)^m = \alpha^{im}\beta^{t^{im}+t^{i(m-1)}+\cdots+t^i} = \alpha^{im}\beta^{V(m)},$$

where $V(m) = t^{im} + t^{i(m-1)} + \cdots + t^i = t^i(t^{i(m-1)} + \cdots + t + 1)$. Let $U(m) = t^{i(m-1)} + \cdots + t + 1$. Now the order of $\beta\alpha^i$ is the least positive integer m such that

$$(\beta\alpha^i)^m = 1 \Leftrightarrow \alpha^{im}\beta^{V(m)} = 1 \Leftrightarrow \alpha^{im} = \beta^{-V(m)} \in \langle\alpha\rangle \cap \langle\beta\rangle = 1 \Leftrightarrow \alpha^{im} = \beta^{V(m)} = 1.$$

Hence m is multiple of $|\alpha^i|$. On the other hand, β has order p and $\beta^{V(m)} = 1 \Leftrightarrow V(m) = 0 \mod p \Leftrightarrow t^i U(m) = 0 \mod p \Leftrightarrow U(m) = 0 \mod p$, since $t \in \mathbb{Z}_p^* = \{1, 2, \ldots, p-1\}$. Now $\ell \mid p-1$ and $t^\ell = 1 \mod p$ and $t^q \neq 1 \mod p$, for any $q < \ell$, that is $t^q \neq 1 \mod p$ for any $q \neq 0 \mod \ell$. Since $i = \theta + j \neq 0 \mod \ell$, then $U(m) = \frac{t^{im}-1}{t^i-1}$, and so

$$U(m) = 0 \mod p \;\Leftrightarrow\; t^{im} = 1 \mod p.$$

Thus, if m is just the order of α^i, that is, if $m = \frac{n}{\gcd(n,i)}$, then im is a multiple of n and so a multiple of ℓ ($n = 0 \mod \ell$), and consequently m also satisfies $\beta^{V(m)} = 1$. Hence $|\beta\alpha^i| = |\alpha^i|$, that is,

$$|a^j b| = |\beta\alpha^{\theta+j}| = |\alpha^{\theta+j}| = \frac{n}{\gcd(n, \theta+j)}.$$

□

5 Number of hypermaps with p hyperfaces

To count the number of regular oriented hypermaps with p (prime) hyperfaces of valency n it suffices to count the different parameters appearing in items (1) and (2) of Theorem 3.2. Let $NH_{(1)}(p,n)$ be the number of regular oriented hypermaps $\mathcal{CM}_{n,p,u,t}$ in item (1) with p hyperfaces of valency n, and $NH_{(2)}(p,n)$ be the number of regular oriented hypermaps $\mathcal{H}^{p,\ell,t,k}_{n,u,v}$ in item (2) with p hyperfaces of valency n. Then

(1) Denote by $U_p(n)$ the subgroup of the units of $\mathbb{Z}_n$ whose elements t satisfy $t^p = 1 \mod n$. Let $\mu(t)$ be the number of solutions u of $(t-1)u = 0 \mod n$. Then

$$NH_{(1)}(p,n) = \sum_{t \in U_p(n)} \mu(t) = \sum_{t \in U_p(n)} \gcd(t-1, n).$$

Now let $NRH_{(1)}(p,n)$ and $NCH_{(1)}(p,n)$ be the number of reflexible and chiral (respectively) regular oriented canonical metacyclic hypermaps $\mathcal{H} = \mathcal{CM}_{n,p,u,t}$ in item (1) with p hyperfaces of valency n. By Theorem 4.1, $\mathcal{H}$ is reflexible if and only if $t^2 = 1 \mod n$. This implies that $t^m = 1 \mod n$ for any even m, and so, combining with $t^p = 1 \mod n$, we get $t = 1 \mod n$. Then

$$NRH_{(1)}(p,n) \;=\; \gcd(0,n) \;=\; n$$

and

$$NCH_{(1)}(p,n) \;=\; \sum_{t\in U_p^*(n)} \gcd(t-1,n)\,,$$

where $U_p^*(n) = \{t \in U_p(n) \mid t \neq 1 \mod n\}$.

(2) Denote by $\wp(t,\ell)$ the number of pairs (u,v) satisfying the equations $u = 0 \mod \ell$, $v-1 = 0 \mod \ell$ and $(H4)$. Since k freely ranges in $\{0,1,\ldots,\ell-1\}$, then

$$NH_{(2)}(p,n) = \sum_{\substack{\ell|\gcd(p-1,n)\\ \ell>1}} \sum_{t\in G_\ell} \sum_{k} \wp(t,\ell) = \sum_{\substack{\ell|\gcd(p-1,n)\\ \ell>1}} \sum_{t\in G_\ell} \ell\wp(t,\ell)\,,$$

where G_ℓ is the set of elements of order ℓ in the cyclic group $\mathbb{Z}_p{}^* = C_{p-1}$. Since p and $t-1$ are coprimes, the number of pairs of solutions $(u, v-1)$ mod n of $(H4)$ is exactly n; so the number $\wp(t,\ell)$ of solutions pairs $(u,v-1) \mod n$ which are multiple of ℓ, where $n = 0 \mod \ell$, of $(H4)$, is the number of solutions pairs $(\frac{u}{\ell}, \frac{v-1}{\ell}) \mod \frac{n}{\ell}$ of

$$\frac{u}{\ell}(t-1) + \frac{v-1}{\ell}p = 0 \mod \frac{n}{\ell}\,,$$

which is exactly $\wp(t,\ell) = \frac{n}{\ell}$. Therefore

$$NH_{(2)}(p,n) = \sum_{\substack{\ell|\gcd(p-1,n)\\ \ell>1}} \sum_{t\in G_\ell} \ell\frac{n}{\ell} \;=\; n \sum_{\substack{\ell|\gcd(p-1,n)\\ \ell>1}} \Phi(\ell)\,,$$

where Φ is the Euler Phi-function. In the special case when p is a Fermat prime, $p-1$ is a power of 2 and so $NH_{(2)}(p,n) = 0$ for n odd. The total number $NH(p,n)$ of regular oriented hypermaps with p (prime) hyperfaces of valency n is then given by:

$$\begin{aligned} NH(p,n) &= NH_{(1)}(p,n) + NH_{(2)}(p,n) \\ &= \sum_{t\in U_p(n)} \gcd(t-1,n) \;+\; \sum_{\substack{\ell|\gcd(p-1,n)\\ \ell>1}} n\Phi(\ell)\,. \end{aligned}$$

Denote by $NRH_{(2)}(p,n)$ and $NCH_{(2)}(p,n)$ the number of reflexible and chiral (respectively) regular oriented hypermaps $\mathcal{H} = \mathcal{H}^{p,\ell,t,k}_{n,u,v}$ in item (2). By Theorem 4.1, $\mathcal{H}$ is reflexible if and only if $t = -1 \mod p$. This is equivalent to $\ell = 2$ (and this implies n even). Hence

$$NRH_{(2)}(p,n) = \begin{cases} 0\,, & \text{if } n \text{ is odd,} \\ n\,, & \text{if } n \text{ is even.} \end{cases}$$

$$NCH_{(2)}(p,n) = n \sum_{\substack{\ell \mid \gcd(p-1,n) \\ \ell > 2}} \Phi(\ell)\,.$$

Note that if n is odd then $\mathcal{H}^{p,\ell,t,k}_{n,u,v}$ is chiral with chirality index p.

Denoting by $NRH(p,n)$ and $NCH(p,n)$ the number of reflexible and chiral regular oriented hypermaps with p (prime) hyperfaces of valency n, then

$$NRH(p,n) = NRH_{(1)}(p,n) + NRH_{(2)}(p,n) = \begin{cases} n\,, & \text{if } n \text{ is odd,} \\ 2n\,, & \text{if } n \text{ is even.} \end{cases}$$

and $NCH(p,n) = NCH_{(1)}(p,n) + NCH_{(2)}(p,n)$.

References

[1] G. V. Belyĭ, Galois extensions of a maximal cyclotomic field, *Izv. Akad. Nauk SSSR Ser. Mat.* **43** (1979), 267–276, 479.

[2] A. Breda, A. Breda D'Azevedo and R. Nedela, Chirality group and chirality index of Coxeter chiral maps, *Ars Combin.* **81** (2006), 147–160.

[3] A. Breda d'Azevedo and M. E. Fernandes, Regular oriented hypermaps up to five hyperfaces, *Acta Univ. M. Belii Ser. Math.* **17** (2010), 21–39.

[4] A. Breda D'Azevedo and M. E. Fernandes, Classification of primer hypermaps with a prime number of hyperfaces, *European J. Combin.* **32** (2011), 233–242, doi:10.1016/j.ejc.2010.09.003, http://dx.doi.org/10.1016/j.ejc.2010.09.003.

[5] A. Breda D'Azevedo, G. Jones, R. Nedela and M. Škoviera, Chirality groups of maps and hypermaps, *J. Algebraic Combin.* **29** (2009), 337–355, doi:10.1007/s10801-008-0138-z, http://dx.doi.org/10.1007/s10801-008-0138-z.

[6] A. Breda D'Azevedo and G. A. Jones, Totally chiral maps and hypermaps of small genus, *J. Algebra* **322** (2009), 3971–3996, doi:10.1016/j.jalgebra.2009.02.017, http://dx.doi.org/10.1016/j.jalgebra.2009.02.017.

[7] A. Breda D'Azevedo and R. Nedela, Chiral hypermaps with few hyperfaces, *Math. Slovaca* **53** (2003), 107–128.

[8] A. Breda d'Azevedo, I. I. Rodrigues and M. E. Fernandes, On chirality groups and regular coverings of regular oriented hypermaps, *Czechoslovak Math. J.* **61(136)** (2011), 1037–1047, doi:10.1007/s10587-011-0046-6, http://dx.doi.org/10.1007/s10587-011-0046-6.

[9] A. J. Breda d'Azevedo, The reflexible hypermaps of characteristic -2, *Math. Slovaca* **47** (1997), 131–153.

[10] R. Cori, *Un code pour les graphes planaires et ses applications*, Société Mathématique de France, Paris, 1975, with an English abstract, Astérisque, No. 27.

[11] D. Corn and D. Singerman, Regular hypermaps, *European J. Combin.* **9** (1988), 337–351, doi:10.1016/S0195-6698(88)80064-7, http://dx.doi.org/10.1016/S0195-6698(88)80064-7.

[12] S.-F. Du, J. H. Kwak and R. Nedela, A classification of regular embeddings of graphs of order a product of two primes, *J. Algebraic Combin.* **19** (2004), 123–141, doi:10.1023/B:JACO.0000023003.69690.18, http://dx.doi.org/10.1023/B:JACO.0000023003.69690.18.

[13] S.-F. Du, J. H. Kwak and R. Nedela, Regular embeddings of complete multipartite graphs, *European J. Combin.* **26** (2005), 505–519, doi:10.1016/j.ejc.2004.02.010, http://dx.doi.org/10.1016/j.ejc.2004.02.010.

[14] A. Grothendieck, Equisse d'un programme (1984), in: P. Lochak and L. Schneps (eds.), *Geometric Galois actions, vol. 2*, Cambridge Univ. Press, Cambridge, volume 242 of *London Math. Soc. Lecture Note Ser.*, pp. 243–284, 1997, doi:10.1017/CBO9780511666124.014, http://dx.doi.org/10.1017/CBO9780511666124.014.

[15] D. L. Johnson, *Topics in the theory of group presentations*, volume 42 of *London Mathematical Society Lecture Note Series*, Cambridge University Press, Cambridge-New York, 1980.

[16] G. Jones and D. Singerman, Maps, hypermaps and triangle groups, in: *The Grothendieck theory of dessins d'enfants (Luminy, 1993)*, Cambridge Univ. Press, Cambridge, volume 200 of *London Math. Soc. Lecture Note Ser.*, pp. 115–145, 1994.

[17] G. A. Jones and D. Singerman, Theory of maps on orientable surfaces, *Proc. London Math. Soc. (3)* **37** (1978), 273–307.

[18] T. R. S. Walsh, Hypermaps versus bipartite maps, *J. Combinatorial Theory Ser. B* **18** (1975), 155–163.

[19] S. Wilson and A. Breda d'Azevedo, Non-orientable maps and hypermaps with few faces, *J. Geom. Graph.* **7** (2003), 173–189.

Author Guidelines

Before submission

Papers should be written in English, prepared in LaTeX, and must be submitted as a PDF file.

The title page of the submissions must contain:

- *Title*. The title must be concise and informative.
- *Author names and affiliations*. For each author add his/her affiliation which should include the full postal address and the country name. If avilable, specify the e-mail address of each author. Clearly indicate who is the corresponding author of the paper.
- *Abstract*. A concise abstract is required. The abstract should state the problem studied and the principal results proven.
- *Keywords*. Please specify 2 to 6 keywords separated by commas.
- *Mathematics Subject Classification*. Include one or more Math. Subj. Class. codes – see http://www.ams.org/msc.

After acceptance

Articles which are accepted for publication must be prepared in LaTeX using class file amcjoucc.cls (which has replaced the deprecated amcjou.cls class file) and the bst file amcjoucc.bst (if you use BibTeX). If you dont use BibTeX, please make sure that all your references are carefully formatted following the examples provided in the sample file.

All files can fe found on-line at:

http://amc-journal.eu/index.php/amc/about/submissions#authorGuidelines

Abstracts: Be concise. As much as possible, please use plain text in your abstract and avoid complicated formulas. Do not include citations in your abstract. All abstracts will be posted on the website in fairly basic HTML, and HTML cant handle complicated formulas. It can barely handle subscripts and greek letters.

Cross-referencing: All numbering of theorems, sections, figures etc. that are referenced later in the paper should be generated using standard LaTeX \label{...} and \ref{...} commands. See the sample file for examples.

Theorems and proofs: The class file has pre-defined environments for theorem-like statements; please use them rather than coding your own. Please use the standard \begin{proof} \end{proof} environment for your proofs.

Spacing and page formatting: Please do not modify the page formatting and do not use \medbreak, \bigbreak, \pagebreak etc. commands to force spacing. In general, please let LaTeX do all of the space formatting via the class file. The layout editors will modify the formatting and spacing as needed for publication.

Figures: Any illustrations included in the paper must be provided in PDF format, or via LaTeX packages which produce embedded graphics, such as TiKz, that compile with PdfLaTeX. (Note, however, that PSTricks is problematic.) Make sure that you use uniform lettering and sizing of the text. If you use other methods to generate your graphics, please provide .pdf versions of the images (or negotiate with the layout editor assigned to your article).

Subscription

Yearly subscription: 150 EUR

Any author or editor that subscribes to the printed edition will receive a complimentary copy of *Ars Mathematica Contemporanea*.

Subscription Order Form

Name: . . . .
E-mail: . . . .
Postal Address: . . . .
. . . .
. . . .
. . . .

I would like to subscribe to receive . . . . . . copies of each issue of *Ars Mathematica Contemporanea* in the year 2016.

I want to renew the order for each subsequent year if not cancelled by e-mail:

☐ Yes ☐ No

Signature: . . . .

Please send the order by mail, by fax or by e-mail.

By mail: Ars Mathematica Contemporanea
UP FAMNIT
Glagoljaška 8
SI-6000 Koper
Slovenia

By fax: +386 5 611 75 71

By e-mail: info@famnit.upr.si

Printed in Slovenia by ******