- École normale supérieure de Lyon -
Laboratoire de l'Informatique du Parallélisme (LIP) - UMR 5668

HABILITATION À DIRIGER DES RECHERCHES

présentée et soutenue publiquement le 27 juin 2017 par

Michaël RAO

Preuves aidées par ordinateur en combinatoire des mots

Devant la commission d'examen formée de :

Srečko	Brlek	*Rapporteur et Examinateur*
Arturo	Carpi	*Rapporteur et Examinateur*
Maxime	Crochemore	*Examinateur*
Juhani	Karhumäki	*Examinateur*
Jarkko	Kari	*Rapporteur*
Mathieu	Sablik	*Examinateur*
Bruno	Salvy	*Examinateur*
Laurent	Vuillon	*Examinateur*

Table des matières

Chapitre 1

Introduction

1.1 Sur l'organisation du mémoire

Je présente dans ce mémoire une sélection de mes travaux depuis ma prise de poste au CNRS. Il est composé de trois parties qui suivent, globalement, un ordre chronologique. Les deux premières parties portent sur la combinatoire des mots, respectivement sur la conjecture de Dejean et sur l'évitabilité des répétitions abéliennes. Toutes les questions abordées sont dans la continuation des travaux de Thue, c'est-à-dire sur l'évitabilité de répétitions dans les mots. La troisième partie expose deux résultats combinatoires que l'on peut rapprocher à des problèmes de mots en dimension deux.

Les chapitres 5 et 6 sont des chapitres de synthèse et écrits pour l'occasion. Les chapitres 3 et 10 correspondent à des manuscrits encore non publiés, et les chapitres restants correspondent à des articles publiés.

La première partie est dédiée à la conjecture de Dejean, ainsi qu'à différentes généralisations. Cette partie s'appuie sur les articles publiés [28, 2, 15], ainsi que sur des résultats encore non publiés, comme en particulier le manuscrit [32].

Cette conjecture, énoncée en 1972 par Françoise Dejean [64], a connu beaucoup d'intérêt en combinatoire des mots, jusqu'à sa démonstration complète en 2009, grâce au travail de plusieurs auteurs. Pour chaque taille d'alphabet, Dejean a conjecturé son *seuil de répétition*, qui est la plus petite puissance fractionnaire qu'un mot infini sur cet alphabet ne peut pas éviter.

Je présente la méthode de ma preuve des derniers cas (chapitre 2), qui est ensuite utilisée pour prouver la conjecture d'Ochem, qui est une généralisation stricte de la conjecture de Dejean (chapitre 3). Cette preuve de la conjecture d'Ochem est également la seconde preuve connue pour la conjecture de Dejean sur les grands alphabets, la première étant due à Carpi [49]. Les chapitres 2 et 3 présentent donc notamment une preuve complète de la conjecture de Dejean pour toutes les tailles d'alphabets. Le chapitre 4 est dédié au *seuil de répétition fini*, c'est-à-dire à un renforcement de la conjecture où l'on veut limiter le nombre de répétitions dont l'exposant est exactement le seuil de répétition limite. Les chapitres 2, 3 et 4 correspondent respectivement aux articles [28], [32] et [2]. Le chapitre 5, concluant cette partie, est une synthèse d'autres questions liées aux mots de Dejean et au seuil de répétition, notamment sur le seuil de répétition généralisé et le taux de croissance des mots de Dejean. J'y expose en particulier des résultats que j'ai obtenus avec Roman Kolpakov [15].

L'idée clef du chapitre 2, qui sera réutilisée dans les chapitres 3 et 4, a été de reprendre le travail de Moulin Ollagnier (qui avait lui-même formalisé le travail de Pansiot), et de l'adapter à des mots morphiques plutôt qu'à des mots purement morphiques. On se retrouve ainsi avec plus de degrés de libertés, qu'il faut savoir gérer plus finement lors des recherches informatisées, mais qui permettent d'avoir des constructions plus petites, et facilement atteignables par nos moyens de calculs. Puis, dans le chapitre 3, l'étude des similarités dans les constructions trouvées par la machine a permis de déduire des mots morphiques fonctionnant pour toutes les tailles d'alphabets.

La deuxième partie du mémoire est dédiée à l'évitabilité des répétitions abéliennes, ainsi que leurs généralisations. Cette partie est basée sur les articles [23, 24, 25, 31].

L'étude de l'évitabilité des répétitions abéliennes débute en 1957 avec des questions d'Erdős, où il demandait s'il est possible d'éviter les carrés abéliens avec un alphabet de 4 lettres. Dekking a montré en 1979 qu'il était possible d'éviter les puissances 4èmes abéliennes sur 2 lettres, et les cubes abéliens sur 3 lettres. La réponse positive à la question d'Erdős, donnée finalement par Keränen en 1992, a nécessité l'utilisation de l'ordinateur afin de trouver une construction.

Le chapitre 6 rappelle les résultats de Dekking et de Keränen, puis présente des généralisations de ces questions : l'évitabilité des grandes répétitions abéliennes, au travers des questions de Mäkelä, et l'évitabilité des répétitions additives, par des questions de Justin, Pirillo et Varricchio. Avec Matthieu Rosenfeld, nous avons obtenu de nouveaux résultats sur les questions de Mäkelä, et sur l'évitabilité des carrés additifs sur $\mathbb{Z}^2$ [24, 25]. Ce chapitre présente ces résultats, sans donner les preuves, qui pourront être trouvées dans le mémoire de thèse de Matthieu Rosenfeld [135].

Récemment, Karhumäki *et al.* [88, 98, 97] ont introduit une généralisation de l'équivalence abélienne, appelée équivalence k-abélienne, qui fait un pont entre l'équivalence abélienne et les répétitions usuelles. Puis Rigo et Salimov [131] ont introduit une autre généralisation, incomparable avec l'équivalence k-abélienne, appelée équivalence k-binomiale. Les questions d'évitabilité se sont naturellement posées pour ces nouvelles notions. Le chapitre 7, basé sur l'article [23], présente un algorithme permettant de décider, sous certaines conditions, si un morphisme évite les puissances k-abéliennes, puis expose une construction de mots infinis binaires (resp. ternaires) interdisant des cubes 2-abéliens (resp. carrés 3-abéliens). Du fait que la méthode utilisée est très similaire, ce chapitre présente également des constructions de mots sans cubes additifs sur des alphabets de taille 3. Le chapitre 8, reprenant l'article [31], expose quant à lui des constructions de mots évitant des carrés et cubes 2-binomiaux. Ces résultats sont, dans tous les cas, optimaux.

Finalement, la troisième partie du manuscrit porte sur deux problèmes combinatoires en dimension deux, et sont significatifs de l'orientation qu'ont pris mes travaux de recherche ces dernières années. Dans le chapitre 9, on démontre une conjecture de Chang de 1992, stipulant la taille minimum d'un ensemble dominant d'une grille (article [10]). Enfin dans le chapitre 10, on présente un jeu de 11 tuiles de Wang apériodique, et on montre qu'il s'agit du plus petit possible (article [11]).

Certains travaux que j'ai menés entre ma thèse et mon habilitation ne sont pas représentés dans ces trois parties. En combinatoire des mots, on peut citer les travaux liés à la densité minimum d'occurrences carrés dans un mot binaire, et dont je vais reparler dans la suite de cette introduction [16, 19]. J'ai également eu des résultats sur l'existence de mots ***self-suffles*** sans carrés [18], ou sur la taille des classes de mots équivalents k-abéliennement [14]. En théorie des graphes, j'ai publié plusieurs articles, souvent liés à des décompositions et dans la continuation de mes travaux de thèse [12, 6, 9, 13, 1, 27, 8]. S'ajoutent à cela de nouvelles bornes sur d'éventuels nombres parfaits impairs [21, 22, 20]. Une liste de mes publications se trouve à la fin du manuscrit.

1.2 Sur l'aide de l'ordinateur

Le fil conducteur de ma recherche est la démonstration de théorèmes en combinatoire à l'aide de l'ordinateur, et beaucoup de ces résultats n'auraient été trouvés si la recherche informatique n'avait pas été correctement maîtrisée. Certaines constructions de mots ont été trouvées après plusieurs jours ou semaines d'exploration, généralement sur des systèmes parallèles. L'exploration de tous les jeux de 11 tuiles de Wang (chapitre 10) a pris environ un an sur une centaine de cœurs. Un bon choix de langage, l'optimisation du code et sa parallélisation était donc souvent des facteurs prédominants pour l'obtention du résultat (tous les programmes qui ont générés des résultats de ce mémoire ont été codés en C++). Je présente ci-après différentes méthodes que j'ai utilisées pour obtenir mes résultats, en particulier en combinatoire des mots. Certaines méthodes n'apparaissent pas clairement dans mes articles, ou dans les chapitres qui vont suivre, car les théorèmes proviennent souvent d'une construction, alors qu'une difficulté souvent cachée, et qu'il n'est pas nécessaire de présenter pour prouver le théorème, est de trouver cette construction. Ce petit inventaire permettra aussi, je l'espère, de voir les liens et similitudes entre les différentes méthodes.

Résultats positifs. Les résultats positifs, c'est-à-dire la démonstration de l'existence d'objets combinatoires possédant certaines propriétés, sont généralement obtenus de manière constructive. Dans le cas des problèmes de combinatoire des mots, ces constructions sont souvent des mots morphiques (autrement dit, des points fixes de morphismes, ou des images de points fixes de morphismes). Il faut, dans ce cas, différencier deux étapes pour l'obtention du résultat. La première étape est généralement de chercher une construction candidate via une exploration par une méthode d'essais et erreurs, souvent agrémentée d'heuristiques. Une fois un candidat trouvé, c'est-à-dire un morphisme dont le point fixe respecte la propriété sur un long préfixe, la deuxième étape est de prouver que ce point fixe, infini, ne contient effectivement aucun facteur interdit. Cette deuxième étape doit généralement aussi être faite par ordinateur, cette fois-ci par un algorithme de décision pour tester la propriété.

Ces deux étapes peuvent être de difficulté variable. Par exemple, dans le cas de mots évitant les puissances abéliennes et k-abéliennes (chapitres 6 et 7), les candidats sont généralement difficiles à trouver, alors que les algorithmes de décision sont plutôt bien connus et compris. Citons que le morphisme de Keränen, répondant à la question d'Erdős, est de taille 85, et est le plus petit possible parmi la classe des morphismes cycliques. C'est aussi pour cette raison que les questions ouvertes de Mäkelä, qui demandent d'éviter les longues puissances abéliennes, posent essentiellement un problème d'ordre calculatoire. D'un autre côté, pour les puissances k-binomiales (chapitre 8), il est facile de trouver des mots morphiques candidats, mais aucun algorithme de décision, ou méthode générique, n'est connu. Ceci explique pourquoi, contrairement à la plupart des autres résultats du manuscrit, les preuves du chapitre 8 sont spécifiques, et non automatisables.

Les recherches des candidats doivent souvent être agrémentées d'heuristiques, qui dépendent, évidemment, fortement du problème. Il y a néanmoins quelques techniques qui se sont montrées fructueuses à plusieurs reprises.

Si on cherche un mot infini évitant un ou des motifs, on peut commencer à chercher un long mot vérifiant la propriété, par le biais d'une recherche exhaustive et un retour sur trace (*backtrack*). Parfois, ce long mot a déjà, ou quasiment, la forme d'un mot morphique. Dans ce cas, une simple analyse du mot nous donnera un morphisme candidat, qu'on pourra ensuite valider, soit à la main, soit par un algorithme de décision pour le problème. Entre autres, c'est cette méthode qui nous a permis de trouver un mot morphique binaire ayant une densité de carrés de $\frac{103}{187}$ [16], qu'on a montré par la suite comme étant le minimum possible [19]. C'est également par cette méthode qu'on a trouvé certains des morphismes du chapitre 4.

S'il ne ressort aucune construction morphique lors de l'analyse d'un long mot dans le langage, l'étape suivante est d'explorer une large classe de morphismes, et de tester si les points fixes ont la propriété voulue. Malheureusement, cette recherche devient rapidement intraitable, même avec la puissance de calcul disponible actuellement. Si aucun morphisme simple ne donne de résultat, on doit réduire l'espace de recherche à des classes de morphismes qu'on suppose être prometteuses. Une possibilité est de limiter l'ensemble des images possibles à des facteurs apparaissant souvent dans les longs mots trouvés aléatoirement avec la propriété (ces mots aléatoires peuvent être trouvés en modifiant la recherche exhaustive par retour sur trace, en fixant un ordre arbitraire sur les lettres de l'alphabet à chaque étape d'extension). On peut également limiter l'ensemble des images à des mots possédant de longs préfixes communs. Ces techniques ont par exemple été fructueuses lors de la recherche du nombre minimum de carrés k-abéliens inévitables dans un mot binaire (théorème 7.5).

Si l'on cherche des morphismes où une ou des lettres peuvent avoir plusieurs images possibles (ce que certains auteurs appellent des substitutions), cela permet en outre de montrer que le langage possède un nombre exponentiel de mots, et d'avoir une borne inférieure sur le taux de croissance du langage (c'est-à-dire $\limsup_{n\to\infty} \sqrt[n]{f(n)}$ où $f(n)$ est le nombre de mots de taille n dans le langage). On peut voir un exemple de cette technique dans la section 7.3.2, où je donne une borne inférieure sur le taux de croissance des mots sans cubes additifs sur 3 lettres. Malheureusement, ces bornes sont généralement assez éloignées des taux de croissance réels des langages.

Il existe également des preuves non constructives d'existence de mots infinis, en utilisant notamment le lemme local de Lovász ou la compression d'entropie. Des méthodes utilisant des séries entières permettent également, non-constructivement, de donner des bornes inférieures sur des taux de croissance. Les résultats du théorème 5.6 illustrent une autre méthode non constructive, due à Kolpakov, qu'on a appliquée aux mots de Dejean. Grossièrement, l'idée de cette méthode est de calculer le nombre des mots du langage $\mathcal{L}^+$ interdisant les répétitions de taille au plus m. Ce taux de croissance peut être calculé exactement, comme expliqué ci-après, en utilisant un graphe de Rauzy. Puis on trouve une borne supérieure sur le nombre de mots du langage $\mathcal{L}^-$, possédant uniquement des répétitions de taille strictement supérieure à m. Ce second nombre étant rapidement négligeable comparativement au premier, on arrive à trouver une borne inférieure pour le nombre de mots dans le langage $\mathcal{L} = \mathcal{L}^+ \setminus \mathcal{L}^-$. On peut noter que la méthode de Kolpakov peut, parfois, donner de bien meilleurs résultats que la méthode des substitutions, et se rapprocher étroitement du taux de croissance du langage, si ce taux de croissance n'est pas trop proche de 1.

Résultats négatifs. Les résultats négatifs, c'est-à-dire de non-existence de mots infinis avec une certaine propriété $\mathcal{P}$, sont généralement prouvés grâce à une recherche exhaustive et un retour sur trace. On sait que si $\mathcal{L}$ est l'ensemble des mots finis avec la propriété héréditaire $\mathcal{P}$, alors l'ensemble $\overline{\mathcal{L}}$ des mots infinis avec la propriété $\mathcal{P}$ est non vide si et seulement si $\mathcal{L}$ est infini. Il suffit donc de parcourir exhaustivement tous les mots de $\mathcal{L}$, et de vérifier que ce langage est fini pour montrer qu'aucun mot infini possède la propriété $\mathcal{P}$. Un exemple dans le manuscrit est le théorème 6.11.

Là encore, des techniques permettent de diminuer considérablement l'espace de recherche. Une première idée est de chercher uniquement à construire un mot minimal via l'ordre lexicographique de $\overline{\mathcal{L}}$. Ainsi, dans l'exploration des mots de $\mathcal{L}$, on coupera la branche courante de l'exploration dès qu'un suffixe propre du mot est plus petit que le préfixe de même taille. De plus, si les lettres de l'alphabet jouent un rôle symétrique, on peut également couper la branche quand un suffixe est plus petit que l'image du préfixe par une permutation des lettres de l'alphabet.

Une autre approche, complémentaire, est d'interdire en plus dans le langage les mots v tels qu'il n'existe pas de $u, u' \in \Sigma^k$ avec $uvu' \in \mathcal{L}$ pour un certain k, qu'on essayera de prendre le plus grand possible. Cela revient à supprimer des sommets dans le graphe de Rauzy associé au langage $\mathcal{L}$. Cette étape peut être répétée tant que des sommets du graphe sont supprimés.

Ces méthodes de recherche exhaustive peuvent aussi être utilisées pour prouver des bornes sur les densités de lettres ou d'occurrences. Mais en général, on obtiendra de meilleurs résultats avec la technique présentée dans la suite, via les matrices de transfert sur les graphes de Rauzy non uniformes.

Matrices de transfert. Enfin, je peux citer dans mes travaux l'utilisation récurrente de méthodes de matrices de transfert, ou des méthodes apparentées. En combinatoire des mots, les méthodes de matrice de transfert permettent d'obtenir différents types des bornes. Calculer le rayon spectral de la matrice d'adjacence du graphe de Rauzy est un moyen classique d'avoir une borne supérieure sur le taux de croissance d'un langage factoriel. C'est en particulier avec cette méthode que les bornes supérieures du théorème 5.6 ont été obtenues.

Dans l'algèbre tropicale $(\min, +)$, l'équivalent du rayon spectral est le *cycle mean* maximum. L'utilisation des matrices de transfert dans l'algèbre tropicale, moins courante, permet d'obtenir des bornes sur les occurrences de facteurs, ou les densités de lettres dans les mots infinis. Cette méthode peut être utilisée pour démontrer la partie négative du théorème 5.1, alors que la partie positive est démontrée par construction. C'est également cette méthode qui a permis de prouver que la densité minimum de carrés dans un mot binaire est $\frac{103}{187}$, en l'utilisant sur la matrice d'adjacence d'un graphe de Rauzy non uniforme [19]. On peut noter que contrairement à la recherche exhaustive présentée précédemment, cette méthode de matrice de transfert dans l'algèbre tropicale permet, parfois, d'obtenir des bornes exactes, autrement dit des bornes atteintes par les constructions. Le chapitre 9 montre une autre utilisation de matrices de transfert dans l'algèbre tropicale, qui a permis de trouver la taille d'un ensemble dominant minimum dans une grille.

1.3 Sur l'orientation de mes recherches

En combinatoire des mots. Parmi les questions sur lesquelles j'ai travaillé en combinatoire des mots, et qui sont restées ouvertes, certaines me captivent particulièrement.

Les premières sont celles liées aux questions de Mäkelä, c'est-à-dire limiter le nombre de répétitions abéliennes inévitables. Malgré nos progrès récents, il semble qu'il reste du chemin à parcourir avant de pouvoir y répondre exactement. Nous avons un schéma de preuve qui pourrait être utilisé, mais des constructions candidates semblent loin d'accès avec nos moyens de calculs. Il faudrait donc trouver des nouvelles méthodes et heuristiques spécifiques au problème. De plus, il n'existe pas, actuellement, de construction simple pour obtenir un mot sans carrés abéliens sur 4 lettres. À ma connaissance, les seules proviennent de Keränen, et de variations dues à Carpi. Nos nouvelles méthodes de preuves permettraient théoriquement d'en trouver d'autres, basées sur des mots morphiques, mais il faudra là encore réussir à dompter l'explosion combinatoire.

D'autres questions sont liées aux méthodes non-constructives, à leurs généralisations et uniformisations. Un cas d'école est la question du *Thue list coloring*, dont la réponse peut sembler triviale, mais qui résiste toujours. Il s'agit de savoir si parmi toute séquence infinie d'ensembles de taille 3, c'est-à-dire un mot infini L sur l'alphabet $\binom{\mathbb{N}}{3}$, on peut toujours extraire un mot w tel que $w[i] \in L[i]$ pour tout i, et tel que w ne possède pas de carré. Une solution constructive peut difficilement être envisagée, du fait même de la formulation du problème. Il semble que le cas le plus contraignant est de choisir des listes identiques, mais on sait dans ce cas que la réponse est vraie, par le résultat de Thue. En utilisant la compression d'entropie, on peut montrer que la réponse est vraie si les listes sont de taille 4, mais la taille 3 semble hors de porté par cette méthode. Un problème est que la compression d'entropie ne considère pas les différentes possibilités d'intersections entre les listes consécutives. À l'opposé, la méthode de Kolpakov est très efficace car elle considère toutes les possibilités de facteurs de petite taille qui peuvent apparaître. Ainsi, cette méthode nécessite beaucoup de calculs, et ne peut pas être utilisée directement sur le problème des listes du fait du nombre démesuré de cas à considérer. Un challenge serait de développer une méthode non-constructive adaptable, en essayant de combiner ces deux méthodes.

Enfin, je ne pourrais terminer cette liste sans parler du fort attrait que j'ai pour les conjectures liées au mot d'Oldenburger, usuellement connu sous le nom de séquence de Kolakoski. J'ai produit de nouvelles bornes sur les densités, de nouvelles conjectures et relations avec des codes bifixes de mots lisses, dont certaines peuvent être trouvées en ligne [26].

Sur les pavages et les problèmes en deux dimensions. J'ai commencé à m'intéresser à des questions de pavages lors de mon affectation au laboratoire J.-V. Poncelet à Moscou avec Thomas Fernique [4]. Plusieurs travaux ont suivi, dont certains se sont montrés fructueux, comme l'obtention du jeu de 11 tuiles de Wang apériodique, et plus récemment la recherche exhaustive de tous les pentagones qui pavent le plan.

En 1996, Kari a proposé un nouveau type de construction de jeux de tuiles de Wang apériodiques. Jusque-là, les jeux de tuiles apériodiques connus étaient construits sur des principes de substitutions. Dans la construction de Kari, le jeu est séparé en deux ensembles A et B, codant respectivement une multiplication (des densités sur les lignes) par des rationnels α et β. La superposition des différentes couches de A et B forment un mot Sturmien, impliquant l'apériodicité.

Un des espoirs de notre recherche exhaustive était de trouver de nouveaux types de constructions. Premièrement, on conjecture qu'il existe une preuve directe de l'apériodicité de notre jeu de 11 tuiles, la preuve actuelle étant assez technique. En effet, il semble que le jeu (après une transformation) peut être séparé en deux ensembles A et B, codant chacun une addition, pour les densités sur les lignes admissibles dans un pavage infini, de $\varphi - 1$ et $\varphi - 2$, où φ est le nombre d'or. Ceci prouverait directement que les superpositions des différentes couches de A et B forment le mot de Fibonacci, qui est apériodique.

Néanmoins, notre jeu de 11 tuiles est assez proche d'un jeu de tuile substitutif. Il serait intéressant de continuer nos recherches, prioritairement sur les jeux de 12 tuiles, pour essayer de trouver des jeux de tuiles apériodiques qui ne suivent pas les constructions classiques (c'est-à-dire ni substitutifs, ni du type de Kari).

Je m'intéresse également au problème de la détermination du taux de croissance (ou de l'entropie) de certains sous-shift de type fini (SFT) en deux dimensions. En une dimension, cela revient à calculer le taux de croissance de langages avec un nombre fini de facteurs interdits, qui est un problème classique ; comme discuté précédemment, le rayon spectral du graphe de Rauzy nous donne la réponse.

Mon intérêt pour ce problème est venu de deux questions *a priori* éloignées. La première est l'étude du nombre de mots de Dejean sur k lettres, quand k tend vers l'infini. Ce problème mène naturellement à des calculs d'entropies de SFT en deux dimensions. Ce sujet sera évoqué en section 5.3.

La deuxième question provient de l'énumération d'ensembles de sommets possédant une certaine propriété sur les graphes (il s'agit d'une branche de recherche en théorie et algorithmique des graphes). Beaucoup de résultats sont des bornes supérieures ou inférieures sur certaines classes de graphes, mais peu de bornes exactes sont connues, ou alors sur des classes de graphes très spécifiques ayant souvent une décomposition en structure arborescente (*treewidth* ou *clique-width* bornées). Les grilles ne possèdent pas de telles décompositions arborescentes, et obtenir des valeurs exactes sur cette classe est donc souvent un problème ardu.

La question m'a été posée sur le nombre d'ensembles dominants dans une grille, alors qu'on venait de résoudre la question sur le taille du plus petit ensemble dominant. Asymptotiquement, cette question peut également se traduire par le taux de croissance d'un SFT en deux dimensions. Une question *a priori* plus simple, mais de la même catégorie, est de trouver le taux de croissance du nombre d'ensembles indépendants dans une grille. Ce taux de croissance peut être vu comme un problème d'entropie dans un SFT en deux dimensions, qui est connu comme le SFT de Fibonacci.

Il s'agit d'un problème qui se retrouve dans d'autres domaines, notamment en mécanique statistique : il y est connu comme l'entropie du modèle *hard square*. Peu de résultats de détermination d'entropies de SFT non triviaux en deux dimensions sont connus, et la plupart des résultats proviennent de ce domaine de la physique. L'un des plus fameux résultats dans ce domaine est celui de Baxter, qui détermine l'entropie du modèle *hard hexagon*, qui correspond aux ensembles indépendants dans une grille triangulaire. Un de mes objectifs est de reproduire et de généraliser ces résultats de mécanique statistique.

Retour aux graphes. Mes travaux sur les graphes, que j'ai commencé dans le cadre de ma thèse, et qui portent majoritairement sur les décompositions de graphes (comme des généralisations de la décomposition modulaire et la *clique-width*) n'ont pas ou peu utilisé l'aide de l'ordinateur.

En tant qu'outil pour aider à prouver des théorèmes, l'ordinateur est généralement moins utilisé dans le domaine de la théorie des graphes que dans les domaines de la combinatoire des mots et des pavages. Ceci est peut-être dû aux faits d'avoir plus de difficultés pour manipuler les graphes dans les langages de programmation, et d'avoir souvent une explosion combinatoire plus difficile à gérer. Il semble aussi que les chercheurs en théorie des graphes sont, en moyenne, plus réticents sur la validité des preuves faites à l'aide de l'ordinateur.

Paradoxalement, un des premiers exemples d'une grosse conjecture prouvée avec l'aide de l'ordinateur est celle du théorème des 4 couleurs, par Appel et Haken en 1977, disant que tout graphe planaire est 4-colorable.

Cette preuve, et beaucoup de preuves de théorèmes semblables sur les graphes planaires, est en deux parties : la première, la phase de *réductions*, montre qu'un contre exemple minimal interdit certaines configurations, et la seconde, le *déchargement*, montre qu'un graphe planaire interdisant toutes ces configurations ne peut pas exister, par une contradiction en partant de la formule d'Euler. La preuve d'Appel et Haken n'a pas été acceptée par tout le monde, du fait d'une part, de l'utilisation de l'ordinateur pour les réductions, et d'autre part, car le déchargement était une étude de cas faite à la main, où différentes erreurs mineures ont été trouvées. Robertson, Sanders, Seymour et Thomas, donnèrent en 1997 une preuve alternative, qui fut ensuite formellement vérifiée en Coq par Gonthier en 2004. Dans cette nouvelle preuve, les deux parties étaient prouvées à l'aide de l'ordinateur.

Beaucoup de théorèmes sur les graphes planaires sont prouvés en utilisant cette même méthode de déchargement. Même si la vérification peut être faite à l'aide de l'ordinateur, cette méthode demande à ce qu'on fournisse une liste de règles de déchargement. Pour le théorème des 4 couleurs, la preuve de 1997 a nécessité une liste de 32 règles de déchargement, trouvées à la main par essais et erreurs. On retrouve une scission semblable à celle discutée au début de la section 1.2 : l'ordinateur peut valider un candidat, ici une liste de règles, mais il faut déjà pouvoir trouver ce candidat. Mais cette recherche de liste candidate n'a jamais été automatisé.

Je m'intéresse depuis quelques années à des méthodes entièrement automatiques afin de prouver qu'un graphe planaire ne peut pas éviter un ensemble de configurations interdites, sans forcément passer par une phase de déchargement.

Un exemple d'un problème qui pourrait être résolu à l'aide d'un nouvel outil de ce style provient à nouveau d'une question d'Erdős : quel est le plus petit k tel que tout graphe planaire sans cycles de taille 4 à k est 3-colorable ? La conjecture de Steinberg, récemment réfutée, proposait $k = 5$ [57]. D'un autre côté, on sait que $k \leq 7$ [44]. On peut aussi imaginer un démonstrateur automatique qui permettrait de prouver des théorèmes du style suivant : pour quelles familles de graphes $\mathcal{H}$, tout graphe planaire ne possédant pas de sous-graphe isomorphe à un graphe dans $\mathcal{H}$ est-il 3-colorable ?

Part I

Dejean's conjecture and beyond

Chapter 2

Last cases of Dejean's conjecture

This first chapter presents the proof of the last cases of Dejean's conjecture [28]. The method used is a generalization of the method of Moulin Ollagnier, and will be used in the next two chapters to show Ochem's conjecture, and to prove the result of the finite repetition threshold over large alphabets.

2.1 Introduction

We use the notation and terminology from Lothaire [111]. We denote by B the set $\{0,1\}$, and by Σ_k the set $\{1,\dots,k\}$ (where $k \geq 2$). Let Σ^* be the set of finite words over the alphabet Σ, and let $\Sigma^+ = \Sigma^* \setminus \{\epsilon\}$ (where ϵ is the empty word). Let Σ^ω be the set of infinite words over Σ. We denote by $w[i]$ the i-th letter of word w, and we denote by $w[i:j]$ (where $i \leq j$) the word $w[i]w[i+1]\dots w[j]$. Let $\mathbb{S}_k$ be the group of permutations of Σ_k, and $\mathrm{Id}_k \in \mathbb{S}_k$ be the identity permutation.

A *square* (resp. a *cube*) in a word is a non-empty factor of the form uu (resp. uuu), and an *overlap* is a factor of the form $xuxux$, where x is a letter and u a (possibly empty) word.

Thue showed that one can avoid squares on ternary words, and overlaps on binary words [143, 142]. (For a translation, see [42].)

Let $\nu_{TM} : \{0,1\}^* \to \{0,1\}^*$ be the morphism such that $\nu_{TM}(0) = 01$ and $\nu_{TM}(1) = 10$. The fixed point of ν_{TM} with first letter 0 is known as the *Thue-Morse word* (or *Prouhet-Thue-Morse word*)

$$w_{TM} = 0110100110010110\dots$$

Thue showed that this word avoids overlaps. In consequence, the Thue-Morse word avoids cubes.

Let $\nu_{TTM} : 0 \to 012, 1 \to 02, 2 \to 1$. The fixed point of ν_{TTM} is known as the *ternary Thue-Morse word*, or the *Hall word*:

$$w_{TTM} = 012021012102012021020121\dots$$

Since $\tau \circ \nu_{TTM} = \nu_{TM} \circ \tau$, with $\tau : 0 \to 011, 1 \to 01, 2 \to 0$, w_{TTM} is also the pre-image of w_{TM} by τ. One can easily show that if w has a square uu, then $\tau(w)$ has an overlap $\tau(u)\tau(u)0$. Thus, since w_{TM} is overlap-free, then w_{TTM} is square-free.

Thue's work gave a complete picture of the avoidability of integral powers in words. Then some authors studied the avoidability of fractional powers, that we call repetitions in the following, and gave rise to a conjecture by Dejean which states the repetition threshold for every alphabet size [64].

A *repetition* in a word w is a pair of words (p,q) such that pq is a factor of w, p is non-empty, and q is a prefix of pq. The *excess* of a repetition (p,q) is $|q|$, and its *exponent* is $\frac{|pq|}{|p|}$. Squares are thus repetitions of exponent 2.

A word is said x-*free* (resp. x^+-free) if it does not contain a repetition of exponent y with $y \geq x$ (resp. $y > x$). For an integer $k \geq 2$, the *repetition threshold* for k letters, denoted by $\mathrm{RT}(k)$, is the infimum over the set of x such that there exists an infinite x-free word over a k-letter alphabet, or equivalently the smallest x such that there exists an infinite x^+-free word over a k-letter alphabet.

Since Thue-Morse word avoids overlaps, *i.e.* is 2^+-free, and than squares are not avoidable over binary words, one have $\mathrm{RT}(2) = 2$. Dejean [64] conjectured that for every $k \geq 2$, $\mathrm{RT}(k) = \mathrm{r}_k$, where:

$$\mathrm{r}_k = \begin{cases} 2 & \text{if } k = 2 \\ \frac{7}{4} & \text{if } k = 3 \\ \frac{7}{5} & \text{if } k = 4 \\ \frac{k}{k-1} & \text{otherwise.} \end{cases}$$

Dejean's conjecture has successively been proved thanks to the work of several authors. The case $k = 3$ was solved by Dejean herself [64]. The case $k = 4$ was solved by Pansiot, introducing the Pansiot's coding [125]. Moulin Ollagnier generalized the idea of Pansiot, and showed the conjecture for $5 \leq k \leq 11$ [118]. Mohammad-Noori and Currie showed the cases $12 \leq k \leq 14$ using Moulin Ollagnier idea on Sturmian words [117]. Carpi showed all the cases over large alphabets $k \geq 33$ [49]. Currie and Rampersad improved Carpi's method for $k \geq 30$ [62] and then $k \geq 27$ [63]. The last cases was independently solved by Rao [28] and Currie and Rampersad [61]. We present in this chapter the proof of Rao which covers the cases $9 \leq k \leq 38$. This method will be used in the next chapter to cover all cases of the stronger conjecture of Ochem.

Dejean showed that $\mathrm{RT}(k) \geq \mathrm{r}_k$ for every $k \geq 2$. This result is immediate for every $k \geq 5$, since $\mathrm{RT}(k) < \mathrm{r}_k$ would imply that $w[i + k] = w[i]$, and the contradiction that the word is periodic. For $k = 3$ and $k = 4$, this can be done by an exhaustive search: there are finitely many $\frac{7}{4}$-free ternary words, and finitely many $\frac{7}{5}$-free words on 4 letters. In order to prove the conjecture for a fixed k, it is thus sufficient to construct an infinite r_k^+-free word over a k-letter alphabet.

Dejean showed that a fixed point of the following morphism is $\frac{7}{4}^+$-free, proving that $\mathrm{RT}(3) = \frac{7}{4}$ [64].

$$h_D : \begin{cases} a \to abcacbcabcbacbcacba \\ b \to bcabacabcacbacabacb \\ c \to cabcbabcabacbabcbac. \end{cases}$$

More generally, this morphism is $\frac{7}{4}^+$-free, that is for every $\frac{7}{4}^+$-free word w, $h_D(w)$ is also $\frac{7}{4}^+$-free. But the method of fixed points has limitations. For example, Brandenburg showed that there is no r_k^+-free morphism over k letters, when $k \geq 4$ [45].

2.2 Pansiot's code and kernel repetitions

Pansiot [125] noticed that if a word on the alphabet Σ_k is $\frac{k-1}{k-2}$-free, then it can be encoded by a binary word. Let $k \geq 3$ and let w be a (possibly infinite) $\frac{k-1}{k-2}$-free word over the alphabet Σ_k, of length at least $k - 1$. Then every factor of length $k - 1$ consists of $k - 1$ different letters. The *Pansiot code* of w is the binary word $P_k(w)$ such that for all $i \in \{1, \ldots, |w| - k + 1\}$ (for all $i \geq 1$ if w is infinite):

$$P_k(w)[i] = \begin{cases} 0 & \text{if } w[i+k-1] = w[i] \\ 1 & \text{if } w[i+k-1] \notin \{w[i], \dots, w[i+k-2]\}. \end{cases}$$

Note that w is uniquely defined by $P_k(w)$ and $w[1 : k-1]$. One can define an inverse operation: for a (possibly infinite) binary word w, $M_k(w)$ is the word on the alphabet Σ_k such that for all $i \in \{1, \dots, |w|+k-1\}$ (for all $i \geq 1$ if w is infinite):

$$M_k(w)[i] = \begin{cases} i & \text{if } i < k, \\ M_k(w)[i-k+1] & \text{if } i \geq k \text{ and } w[i-k+1] = 0, \\ \alpha & \text{otherwise,} \end{cases}$$

where $\{\alpha\} = \Sigma_k \setminus \{M_k(w)[i-k+1], \dots, M_k(w)[i-1]\}$. Note that if $w[i] = i$ for all $i < k$, then $M_k(P_k(w)) = w$.

Let h_P be the following morphism:

$$h_P : \begin{cases} 0 \to 101101 \\ 1 \to 10. \end{cases}$$

Pansiot showed that $M_4(h_P^\infty(1))$ is $\frac{7}{5}^+$-free, proving Dejean's conjecture for 4 letter alphabets [125].

Moulin Ollagnier showed that Pansiot's coding can also be viewed by the way of an action on the symmetric group $\mathbb{S}_k$ [118]. Let Ψ be the morphism between the free monoid B^* and $\mathbb{S}_k$ such that $\Psi(0) = \sigma_0$ and $\Psi(1) = \sigma_1$, where:

$$\sigma_0 = \begin{bmatrix} 1 & 2 & \cdots & k-2 & k-1 & k \\ 2 & 3 & \cdots & k-1 & 1 & k \end{bmatrix} \quad \text{and} \quad \sigma_1 = \begin{bmatrix} 1 & 2 & \cdots & k-2 & k-1 & k \\ 2 & 3 & \cdots & k-1 & k & 1 \end{bmatrix}.$$

One can easily show that for all $i \geq 0$ and $1 \leq j \leq k-1$, $M_k(w)[i+j] = \Psi(w[1 : i])(j)$.

Suppose that an infinite word w validates Dejean's conjecture for k letters with $k \geq 3$. We can suppose without loss of generality that $w[i] = i$ for all $i < k$, and that w can be encoded by its Pansiot's code, since $\mathrm{r}_k < \frac{k-1}{k-2}$. Thus Dejean's conjecture is true for a $k \geq 3$ if and only if there exists an infinite binary word w such that $M_k(w)$ is r_k^+-free.

Let w be a (possibly infinite) word on the alphabet Σ. A *Φ-kernel repetition* in w (where $\Phi : \Sigma^* \to \mathbb{S}_k$ is a morphism) is a pair (p, q) such that (p, q) is a repetition in w and $\Phi(p) = \mathrm{Id}_k$.

We fix a $k \geq 3$. We say that a repetition is *forbidden* if $\frac{|pq|}{|p|} > \mathrm{r}_k$. A repetition is a *short repetition* if $|q| < k-1$, otherwise it is called a *kernel repetition*. Note that if (p, q) is a forbidden short repetition, then $|p| < \frac{k-2}{\mathrm{r}_k - 1}$. Moulin Ollagnier [118] showed that:

Proposition 2.1 ([118]). *Let $k \geq 3$ and w be a binary word. Then $M_k(w)$ has a kernel repetition (p, q) if and only if w has a Ψ-kernel repetition (p', q') with $|p'| = |p|$ and $p'q' = P_k(pq)$. (Note that $|q'| = |q| - k + 1$.)*

Moulin Ollagnier gave necessary and decidable conditions for a morphism h to have fixed points without forbidden Ψ-kernel repetition, and gave morphisms for $3 \leq k \leq 11$ which validate Dejean's conjecture. His ideas have then been adapted to morphic Sturmian words by Mohammad-Noori and Currie [117], proving conjecture for $7 \leq k \leq 14$. We extend here Moulin Ollagnier's ideas to morphic words, and in particular words which are the image by a morphism of the Thue-Morse word.

2.3 Preliminary results

We first introduce some notations and prove technical results, which are simple adaptations of Moulin Ollagnier's ideas. Throughout this section, $f : \{x, y\}^* \to \{z, t\}^*$ denotes a morphism (possibly with $\{x, y\} = \{z, t\}$), such that:

(LL) the last letters of $f(x)$ and $f(y)$ differ, and

(PC) $\{f(x), f(y)\}$ is a prefix-code, *i.e.* $f(x)$ is not a prefix of $f(y)$, and $f(y)$ is not a prefix of $f(x)$.

Let L be the largest common prefix of $f(x)$ and $f(y)$, and let $\ell = |L|$.

Definition 2.2 (Interpretation, Markable)**.** Let w be an infinite word on the alphabet $\{x, y\}$. Let v be a non-empty factor of $f(w)$. An (f, w)-*interpretation* of v is a triplet (b, u, e) such that:

- u is a non-empty factor of w,
- the *beginning* b is a non-empty suffix of $f(u[1])$,
- the *end* e is a non-empty prefix of $f(u[|u|])$,
- $e'vb' = f(u)$, where $e'b = f(u[1])$ and $eb' = f(u[|u|])$.

A word u is (f, w)-*markable* if all its (f, w)-interpretations have the same beginning.

Proposition 2.3. *Let v be a factor of $f(w)$ and v' be a factor of v. If v' is (f, w)-markable, then v is (f, w)-markable.*

Proof. It suffices to show that if $v[1 : |v| - 1]$ or $v[2 : |v|]$ is (f, w)-markable, then v is (f, w)-markable. Obviously, if $v[1 : |v| - 1]$ is (f, w)-markable, then v is (f, w)-markable.

Suppose that $v[2 : |v|]$ is (f, w)-markable, and let b be the unique beginning of all its (f, w)-interpretations. Let $\gamma \in \{x, y\}$ such that b is a suffix of $f(\gamma)$. By definition, γ exists, and by condition (LL), γ is unique. Let e' be such that $e'b = f(\gamma)$. If e' is non-empty, then $v[1] = e'[|e'|]$, and $v[1]b$ is the only possible beginning of an (f, w)-interpretation of v. Otherwise, the only possible beginning of an (f, w)-interpretation of v is $v[1]$. $\square$

Proposition 2.4. *Let v be an (f, w)-markable factor of $f(w)$. Then v has at most two (f, w)-interpretations. Moreover if v has two different (f, w)-interpretations (b, u_1, e_1) and (b, u_2, e_2), then $e_1 = e_2$, e_1 is a prefix of L, $|u_1| = |u_2|$ and u_1 and u_2 only differ by their last letter.*

Proof. Suppose that v has two different (f,w)-interpretations (b,u_1,e_1) and (b,u_2,e_2). By condition (LL), there is a unique $\alpha \in \{x,y\}$ such that b is a suffix of $f(\alpha)$. Thus $u_1[1] = u_2[1] = \alpha$. Let e' be such that $e'b = f(\alpha)$. Since $\{f(x), f(y)\}$ is a prefix-code (condition (PC)), there is a unique u' such that $f(u')$ is a prefix of $e'v$ and such that for every $\beta \in \{x,y\}$, $f(u'\beta)$ is not a prefix of $e'v$. Obviously, u' is a prefix of u_1 and u_2. Now let e'' be such that $f(u')e'' = e'v$.

If e'' is empty, then $u_1 = u_2 = u'$ and $e_1 = e_2 = f(u'[|u'|])$, contradiction. Thus $e_1 = e_2 = e''$ and $u_1 \neq u_2$. Moreover, since $e'v$ is a prefix of $f(u_1)$ and $f(u_2)$, u_1 and u_2 have length $|u'|+1$, thus e'' is a prefix of both $f(x)$ and $f(y)$, and e'' is a prefix of L. This implies also that there is no other (f,w)-interpretation of v, since every (f,w)-interpretation is of the form $(b, u'\gamma, e'')$ with $\gamma \in \{x,y\}$. □

A repetition (p,q) of w *extends* a repetition (p',q') if $|p| = |p'|$ and $p'q'$ is a proper factor of pq. A repetition (p,q) in w is *maximal* if there is no repetition (p',q') in w such that (p',q') extends (p,q). Note that if (p',q') extends (p,q), then (p,q) is a Φ-kernel repetition if and only if (p',q') is a Φ-kernel repetition (where $\Phi : \Sigma \to \mathbb{S}_k$ is a morphism), since that in this case p and p' are two conjugate words, and thus $\Phi(p) = \mathrm{Id}_k$ if and only if $\Phi(p') = \mathrm{Id}_k$.

Lemma 2.5. *Let $\Phi : \{z,t\}^* \to \mathbb{S}_k$ and $\Phi' : \{x,y\}^* \to \mathbb{S}_k$ be two morphisms such that:*

(CO) there is $\sigma \in \mathbb{S}_k$ such that for every $\alpha \in \{x,y\}$, $\Phi(f(\alpha)) = \sigma \cdot \Phi'(\alpha) \cdot \sigma^{-1}$.

Let (p,q) be a Φ-kernel repetition in $f(w)$ such that q is (f,w)-markable. Then w has a Φ'-kernel repetition (p',q') with $|p| = |f(p')|$ and $|q| \leq |f(q')| + \ell$.

Proof.

Claim 2.6. Suppose that (p,q) is not maximal, and suppose that there is no maximal Φ-kernel repetition extending (p,q). Then for every integer n, there is a Φ'-kernel repetition (p',q') of w such that $|p| = |f(p')|$ and the exponent of (p',q') is at least n.

Proof. There is an infinite sequence $((p,q) = (p_0,q_0), (p_1,q_1), \ldots)$ of repetitions of $f(w)$ such that (p_{i+1}, q_{i+1}) extends (p_i,q_i) for every $i \geq 0$. Clearly, for every $n \geq 1$, there is an $i_n \in \mathbb{N}$ such that p^n is a factor of $p_{i_n}q_{i_n}$, and *a fortiori* of $f(w)$.

Since q is (f,w)-markable, there is a $n_0 > 0$ such that p^{n_0} is (f,w)-markable. Note that for every conjugate word $\widetilde{p}$ of p, $\widetilde{p}^{n_0+1}$ is an (f,w)-markable factor of $f(w)$, and for every $n > 0$, $(\widetilde{p}, \widetilde{p}^n)$ is a Φ-kernel repetition of $f(w)$.

Let $\widetilde{p}$ be a conjugate word of p such that every (f,w)-interpretation of $\widetilde{p}^{n_0+1}$ has a beginning $b \in \{f(x), f(y)\}$. Let (b,u,e) be an (f,w)-interpretation of $\widetilde{p}^{n_0+2}$, and let p' be the prefix of u such that $f(p') = \widetilde{p}$. Such a prefix exists, since $\widetilde{p}^{n_0+1}$ is (f,w)-markable and has $b = f(u[1])$ as beginning.

By condition (CO), $\Phi'(p') = \Phi(\widetilde{p}) = \Phi(p) = \mathrm{Id}_k$. Thus for every $n > 0$, (p', p'^n) is a Φ'-kernel repetition of w, and $|f(p')| = |\widetilde{p}| = |p|$. □

By Claim 2.6, if (p,q) is not maximal and there is no maximal Φ-kernel repetition extending (p,q), then w has a Φ'-kernel repetition (p',q') with $|p| = |f(p')|$ and $|q| \leq |f(q')| + \ell$.

Thus we can now suppose without loss of generality that (p, q) is maximal (otherwise, replace (p, q) by a maximal repetition which extends (p, q)). Let (b, u, e) be an (f, w)-interpretation of q. By Proposition 2.4, all (f, w)-interpretations of q has beginning b and end e. By Proposition 2.3, pq is markable. Since q is a prefix and a suffix of pq, the beginning (resp. end) of an (f, w)-interpretation of pq is b (resp. e). Let (b, v, e) be a (f, w)-interpretation of pq. By maximality of (p, q), $b = f(v[1])$.

If q has only one (f, w)-interpretation, then u is a prefix and a suffix of v, and by maximality of (p, q), $e = f(v[|v|])$. Let p' be such that $p'u = v$, and let $q' = u$. Clearly $f(q') = q$, $f(p'q') = pq$ and $f(p') = p$. By condition (CO), $\Phi'(p') = \Phi(p) = \mathrm{Id}_k$, thus (p', q') is a Φ'-kernel repetition. (Note that in this case, $\ell = 0$ by maximality of (p, q).)

Otherwise, by Proposition 2.4, let (b, u_1, e) and (b, u_2, e) be the only two possible (f, w)-interpretations of q, and let $u' = u_1[1 : |u_1| - 1]$. Then u' is a prefix and a suffix of $v[1 : |v| - 1]$. Let p' be such that $p'u' = v[1 : |v| - 1]$. Clearly $f(p') = p$, and by Proposition 2.4, $|e| \leq \ell$, thus $|q| \leq |f(q')| + \ell$, with $q' = u'$. By condition (CO), $\Phi'(p') = \Phi(p) = \mathrm{Id}_k$, thus (p', q') is a Φ'-kernel repetition. □

2.4 Images of the Thue-Morse word

We recall that ν_{TM} denotes the Thue-Morse morphism, and that w_{TM} is the fixed point of ν_{TM} starting by 0.

Let h be a morphism from $\{a, b\}^*$ into $\{0, 1\}^*$. Let $\sigma_a = \Psi(h(a))$ and $\sigma_b = \Psi(h(b))$ (we recall that Ψ is the morphism from the free monoid $\{0, 1\}^*$ into $\mathbb{S}_k$ such that $\Psi(0) = \sigma_0$ and $\Psi(1) = \sigma_1$). We suppose that h respects the following preliminary conditions:

(A) h is uniform (*i.e.* $|h(a)| = |h(b)|$),

(B) the last letters of $h(a)$ and $h(b)$ differ,

(C) $\{h(a), h(b)\}$ is *comma-free*, that is for every $(x, y, z) \in \{a, b\}^3$, there is no $w, w' \in \{0, 1\}^+$ such that $h(yz) = w \cdot h(x) \cdot w'$, and

(D) there is $\sigma \in \mathbb{S}_k$ such that $\sigma_a\sigma_b = \sigma \cdot \sigma_a \cdot \sigma^{-1}$ and $\sigma_b\sigma_a = \sigma \cdot \sigma_b \cdot \sigma^{-1}$.

Note that ν_{TM} and h respect conditions (LL) and (PC) of Section 2.3. By Proposition 2.1, $M_k(h(w_{TM}))$ has no forbidden repetition if and only if:

(S) $M_k(h(w_{TM}))$ has no forbidden short repetition, and

(K) for every Ψ-kernel repetition (p, q) in $h(w_{TM})$, we have $\frac{|pq|+k-1}{|p|} \leq \mathrm{r}_k$.

Let L be the largest common prefix of $h(a)$ and $h(b)$, let $\ell = |L|$ and let $s = |h(a)| = |h(b)|$. Let Ψ' be the morphism from the free monoid $\{a, b\}^*$ to $\mathbb{S}_k$ such that $\Psi'(a) = \sigma_a$ and $\Psi'(b) = \sigma_b$.

Lemma 2.7. *Let u be a factor of $h(w_{TM})$ with $|u| \geq 2s-1$. Then u is (h, w_{TM})-markable.*

Proof. Obviously $h(a)$ and $h(b)$ are (h, w_{TM})-markable since $\{h(a), h(b)\}$ is comma-free. Thus Lemma 2.7 follows from Proposition 2.3 and from the fact that if $|u| \geq 2s - 1$, then u contains either $h(a)$ or $h(b)$ as factor. □

As corollary of Lemma 2.7 and Lemma 2.5 with $f = h$, $\{x, y\} = \{a, b\}$, $\{z, t\} = \{0, 1\}$, $\Phi = \Psi$ and $\Phi' = \Psi'$, we obtain (condition (CO) is trivially fulfilled with $\sigma = \mathrm{Id}_k$) :

Corollary 2.8. *Let (p, q) be a Ψ-kernel repetition in $h(w_{TM})$ such that $|q| \geq 2s-1$. Then there is a Ψ'-kernel repetition (p', q') in w_{TM} such that $|p| = |h(p')|$ and $|q| \leq |h(q')| + \ell$.*

Obviously, if (p', q') is a Ψ'-kernel repetition in w_{TM}, then $h(w_{TM})$ has a Ψ-kernel repetition (p, q) with $p = h(p')$ and $q = h(q')L$. Thus $M_k(h(w_{TM}))$ has no forbidden kernel repetition if and only if:

(SK) for every Ψ-kernel repetition (p, q) in $h(w_{TM})$ with $|q| < 2s - 1$, we have $\frac{|pq|+k-1}{|p|} \leq \mathrm{r}_k$, and

(LK) for every Ψ'-kernel repetition (p', q') in w_{TM}, we have $\frac{s|p'q'|+\ell+k-1}{s|p'|} \leq \mathrm{r}_k$.

Lemma 2.9. *Let w be a factor of w_{TM} of length at least 4. Then w is (ν_{TM}, w_{TM})-markable.*

Proof. This follows from the fact that w_{TM} avoids aaa, bbb, $ababa$ and $babab$, and thus aa, bb, $abab$ and $baba$ have unique (ν_{TM}, w_{TM})-interpretations and are (ν_{TM}, w_{TM})-markable. ☐

As corollary of Lemma 2.9 and Lemma 2.5 with $f = \nu_{TM}$, $\{x, y\} = \{z, t\} = \{a, b\}$, $\Phi = \Phi' = \Psi'$, we obtain (condition (CO) is fulfilled by condition (D)):

Corollary 2.10. *Let (p', q') be a Ψ'-kernel repetition in w_{TM} with $|q'| \geq 4$. Then there is a Ψ'-kernel repetition (p'', q'') in w_{TM} with $|p'| = 2 \cdot |p''|$ and $|q'| \leq 2 \cdot |q''|$.*

Note that in this case, $\frac{s|p''q''|+\ell+k-1}{s|p''|} \geq \frac{s|p'q'|+\ell+k-1}{s|p'|}$. Thus by Corollary 2.10, condition (LK) is equivalent to:

(LK') for every Ψ'-kernel repetition (p'', q'') in w_{TM} with $|q''| \leq 3$, we have $\frac{s|p''q''|+\ell+k-1}{s|p''|} \leq \mathrm{r}_k$.

2.5 Decidability and results

Let $k \geq 3$, let $h : \{a, b\}^* \to \Sigma_k^*$ be a morphism which respects conditions (A-D), and let $\Psi' : \{a, b\}^* \to \mathbb{S}_k$ be the morphism such that $\Psi'(a) = \sigma_a = \Psi(h(a))$ and $\Psi'(b) = \sigma_b = \Psi(h(b))$. If Ψ' respects condition (LK'), then by Corollary 2.10, w_{TM} has no Ψ'-kernel repetition (p', q') with $\frac{s|p'q'|+\ell+k-1}{s|p'|} > \mathrm{r}_k$, where $s = |h(a)| = |h(b)|$ and ℓ is the size of the largest common prefix of $h(a)$ and $h(b)$. Moreover, if h respects (SK), then by Corollary 2.8, $h(w_{TM})$ has no Ψ-kernel repetition (p, q) with $\frac{|pq|+k-1}{|p|} > \mathrm{r}_k$, that is by Proposition 2.1, $M_k(h(w_{TM}))$ has no forbidden kernel repetition. Finally if h also respects (S), then $M_k(h(w_{TM}))$ has no forbidden repetition. To summarize, if h and Ψ' respect conditions (A-D), (S), (SK) and (LK'), then $M_k(h(w_{TM}))$ validates Dejean's conjecture for k letters.

Obviously, conditions (A-D) are decidable. Condition (S) is decidable since short forbidden repetitions have length less than $\frac{\mathrm{r}_k \times (k-2)}{\mathrm{r}_k - 1}$. It is sufficient to check every factor of size at most $\frac{\mathrm{r}_k \times (k-2)}{\mathrm{r}_k - 1}$ in $M_k(h(w_{TM}))$. Similarly, if (p, q) is a Ψ-kernel repetition of $h(w_{TM})$ with $|q| \leq 2s - 2$ and $\frac{|pq|+k-1}{|p|} > \mathrm{r}_k$, then $|pq| < \frac{\mathrm{r}_k \times (2s-2)+k-1}{\mathrm{r}_k - 1}$. Condition (SK) is thus decidable. Finally, if (p', q') is a Ψ'-kernel repetition of w_{TM} with $\frac{s|p'q'|+\ell+k-1}{s|p'|} > \mathrm{r}_k$ and $|q'| \leq 3$, then $|p'| < \frac{k-1+\ell+3s}{s \times (\mathrm{r}_k - 1)}$. Thus condition (LK') is decidable.

For every $k \in \{4, 8, \ldots, 38\}$, we found a morphism which respects conditions (A-D), (S), (SK) and (LK'). This proves that Dejean's conjecture holds for $8 \leq k \leq 38$. Following, we give an example for $k = 18$. Morphisms h_k^+ and h_k^-, for $k \in \{9, \ldots 38\}$, prove Dejean's conjecture as well Ochem's conjecture on k letters, as we will see in the next chapter. We will also see that the method of this chapter can also gives Pansiot code of Dejean words starting from other morphisms that ν_{TM}.

For every $k \in \{2, 3, 5, 6, 7\}$ and for every $\sigma_a, \sigma_b \in \mathbb{S}_k$ such that $\sigma_a \cdot \sigma_b = \sigma \cdot \sigma_a \cdot \sigma^{-1}$ and $\sigma_b \cdot \sigma_a = \sigma \cdot \sigma_b \cdot \sigma^{-1}$ for a $\sigma \in \mathbb{S}_k$, the word w_{TM} has a Ψ'-kernel repetition (p, q) with $\frac{|pq|}{|p|} \geq \mathrm{r}_k$, thus the technique presented in Section 2.4 (image of the Thue-Morse word) is not applicable in these cases. Nevertheless, we will show in Section 3.4 that other morphic words will work.

Example: $k = 18$**.** Let h_{18} be the morphism such that:

$$h_{18} : \begin{cases} a \rightarrow 10101101010110101101010110110101011010110 \\ b \rightarrow 10101011010110101101011010110101011010101. \end{cases}$$

Then $s = |h_{18}(a)| = |h_{18}(b)| = 41$, $L = 10101$, $\ell = 5$, and we have:

$$\begin{cases} \sigma_a = \Psi(h_{18}(a)) = \left[\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 \\ 7 & 9 & 6 & 8 & 10 & 11 & 14 & 12 & 16 & 13 & 15 & 2 & 17 & 4 & 18 & 1 & 3 & 5 \end{smallmatrix}\right] \\ \sigma_b = \Psi(h_{18}(b)) = \left[\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 \\ 6 & 9 & 7 & 8 & 10 & 11 & 14 & 12 & 16 & 13 & 15 & 2 & 17 & 4 & 18 & 3 & 1 & 5 \end{smallmatrix}\right]. \end{cases}$$

Obviously, h_{18} respects (A) and (B). Moreover, it is not hard to show that h_{18} respects condition (C). Finally, h_{18} respects (D) with:

$$\sigma = \left[\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 \\ 1 & 4 & 3 & 10 & 6 & 14 & 11 & 17 & 12 & 15 & 8 & 7 & 5 & 18 & 2 & 9 & 13 & 16 \end{smallmatrix}\right].$$

Now if (p, q) is a forbidden short repetition in $M_{18}(h_{18}(w_{TM}))$ (*i.e.* $|q| \leq 16$), then $|pq| < 288$. If (p, q) is a Ψ-kernel repetition in $h_{18}(w_{TM})$ with $|q| < 2 \times 41 - 1$ and $\frac{|pq|+k-1}{|p|} > \frac{18}{17}$, then $|pq| < 1729$, and thus $M_{18}(h_{18}(w_{TM}))$ has a forbidden repetition of length less than 1746. Thus to check conditions (S) and (SK), it is sufficient to check whether $M_{18}(h_{18}(\nu_{TM}^8(a)))$ contains no forbidden repetition, since $\nu_{TM}^8(a)$ contains all factors of w_{TM} of length at most 65. Finally, if (p', q') is a Ψ'-kernel repetition in w_{TM} with $|q'| \leq 3$ and $\frac{s|p'q'|+\ell+k-1}{s|p'|} \leq \frac{18}{17}$, then $|p'| \leq 60$. To check (LK'), it is thus sufficient to check every Ψ'-kernel repetition in $\nu_{TM}^8(a)$.

Morphisms for $k \in \{4,8\}$

$$h_4\colon \begin{cases} a \to 10110101011011010110110101011011010101101 \\ \phantom{a \to{}} 1101010110101011011010110110101011011010 \\ b \to 1011010101101101011011010101101010110110 \\ \phantom{b \to{}} 1010110110101011011010110110101011010101 \end{cases}$$

$$h_8\colon \begin{cases} a \to 10110110110101010101101010101101 \\ b \to 10110101010101101101101101101010 \end{cases}$$

Morphisms h_k^+ for $9 \le k \le 38$

$$h_9^+\colon \begin{cases} a \to 1\underline{0}1011011\underline{0}1101011\underline{0}1101101\underline{0}1011011\underline{0}101010 \\ b \to 1\underline{0}1101101\underline{0}1010101\underline{0}1011011\underline{0}1011011\underline{0}110101 \end{cases}$$

$$h_{10}^+\colon \begin{cases} a \to 1\underline{0}10101011\underline{0}10101101\underline{0}11011011\underline{0}11011011\underline{0}10101011\underline{0}1010110 \\ b \to 1\underline{0}10110101\underline{0}10101011\underline{0}11011011\underline{0}11011011\underline{0}11011011\underline{0}1101101 \end{cases}$$

$$h_{11}^+\colon \begin{cases} a \to 10101\underline{0}101101101\underline{0}110101011\underline{0}101101101\underline{0}101010101\underline{0}101010101\underline{0}1101 \\ b \to 10101\underline{0}101101101\underline{0}110101011\underline{0}101011011\underline{0}101011011\underline{0}101010101\underline{0}1010 \end{cases}$$

$$h_{12}^+\colon \begin{cases} a \to 1\underline{0}1101101101\underline{0}1101010101\underline{0}1101011011\underline{0}1010110101\underline{0}101010110 \\ b \to 1\underline{0}1010101101\underline{0}1101101101\underline{0}1011010101\underline{0}1010110101\underline{0}110110101 \end{cases}$$

$$h_{13}^+\colon \begin{cases} a \to 1\underline{0}11010101101\underline{0}10110110101\underline{0}1011011010 \\ \phantom{a \to{}} 1\underline{0}11010101101\underline{0}10110101011\underline{0}1011010101 \\ b \to 1\underline{0}11010101011\underline{0}11010101011\underline{0}1011010110 \\ \phantom{b \to{}} 1\underline{0}11010101101\underline{0}10110110101\underline{0}1011011010 \end{cases}$$

$$h_{14}^+\colon \begin{cases} a \to 1\underline{0}110101101011\underline{0}110101101101\underline{0}110110101011\underline{0}110101011011\underline{0}11010110101 \\ b \to 1\underline{0}110110101101\underline{0}110110101011\underline{0}110101101101\underline{0}110110101101\underline{0}11011011010 \end{cases}$$

$$h_{15}^+\colon \begin{cases} a \to 1\underline{0}1010101101011\underline{0}1010101101101\underline{0}101011010101 \\ \phantom{a \to{}} 1\underline{0}1010101011011\underline{0}101010110101 \\ b \to 1\underline{0}1010110101101\underline{0}1010110101011\underline{0}101010110110 \\ \phantom{b \to{}} 1\underline{0}1010110101101\underline{0}101011011010 \end{cases}$$

$$h_{16}^+\colon \begin{cases} a \to 1\underline{0}11010101101011\underline{0}10101101011011\underline{0}11011010110101\underline{0}1101010110101 \\ b \to 1\underline{0}11010101011011\underline{0}10101101101011\underline{0}11011010101101\underline{0}1101010110110 \end{cases}$$

$$h_{17}^+\colon \begin{cases} a \to 1\underline{0}101101101011011\underline{0}110101011010101\underline{0}101101011011011\underline{0}11010110101010 \\ b \to 1\underline{0}110101011011011\underline{0}101101101010101\underline{0}110101101010101\underline{0}11010101101101 \end{cases}$$

$$h_{18}^+\colon \begin{cases} a \to 1\underline{0}1011011011011011\underline{0}1010101011010101\underline{0}1010101010101101\underline{0}110110110110110 \\ b \to 1\underline{0}1011011011011011\underline{0}1010101010110101\underline{0}1010101011010101\underline{0}110110110101101 \end{cases}$$

$$h_{19}^+\colon \begin{cases} a \to 1\underline{0}10110101101011011\underline{0}11010101101010101\underline{0}1011011010101101 \\ b \to 1\underline{0}10110110101011011\underline{0}11010101101011011\underline{0}1101011010101010 \end{cases}$$

$$h_{20}^+\colon \begin{cases} a \to 1\underline{0}110110110101101101\underline{0}11011010110110110 \\ \phantom{a \to{}} 1\underline{0}110110101010101011\underline{0}11010101101010110 \\ b \to 1\underline{0}110110110101101011\underline{0}11011010110110101 \\ \phantom{b \to{}} 1\underline{0}110110101010101101\underline{0}11010101101010101 \end{cases}$$

$$h_{21}^+\colon \begin{cases} a \to 1\underline{0}1101010110110101101\underline{0}1010110101010101011\underline{0}101101010101010110 \\ b \to 1\underline{0}1101010110110101011\underline{0}1010110101010101101\underline{0}101101010101010101 \end{cases}$$

$$h_{22}^+\colon \begin{cases} a \to 1\underline{0}11011010101101101011\underline{0}11011010110101101011\underline{0}1010101101011010110 \\ b \to 1\underline{0}11011010101101101101\underline{0}11011010110101101101\underline{0}1010101101011010101 \end{cases}$$

$$h_{23}^+\colon \begin{cases} a \to 1\underline{0}110110110101010101011\underline{0}101101010110110101011\underline{0}10101101010101010110 \\ b \to 1\underline{0}110110110101010101101\underline{0}101101010110110101101\underline{0}10101101010101010101 \end{cases}$$

$$h_{24}^+\colon \begin{cases} a \to 1\underline{0}1011010110110101101011\underline{0}101011011011011010110 \\ b \to 1\underline{0}1011010110110101101101\underline{0}101011011011011010101 \end{cases}$$

$$h_{25}^+\colon \begin{cases} a \to 1\underline{0}10110101101101010101011\underline{0}110101101011010101010110 \\ b \to 1\underline{0}10110101101101010101101\underline{0}110101101011010101010101 \end{cases}$$

$$h_{26}^+\colon \begin{cases} a \to 1\underline{0}101101101011011010101011\underline{0}11010110101101101010110 \\ b \to 1\underline{0}101101101011011010101101\underline{0}11010110101101101010101 \end{cases}$$

$$h_{27}^+\colon \begin{cases} a \to 1\underline{0}1011010110101101010101011\underline{0}11010101101011010101010110 \\ b \to 1\underline{0}1011010110101101010101101\underline{0}11010101101011010101010101 \end{cases}$$

$$h_{28}^{+}\colon \begin{cases} a \to 1\underline{0}10110101011010101010101011\underline{0}1101010101101010101010110 \\ b \to 1\underline{0}1011010101101010101010101101\underline{0}1101010101101010101010101 \end{cases}$$

$$h_{29}^{+}\colon \begin{cases} a \to 1\underline{0}101010110101010101010101011\underline{0}10101010110101010101010110 \\ b \to 1\underline{0}1010101101010101010101011 01\underline{0}10101010110101010101010101 \end{cases}$$

$$h_{30}^{+}\colon \begin{cases} a \to 1\underline{0}10101011011011011010101010101 1\underline{0}1101101101010101101010110 \\ b \to 1\underline{0}1010101101101101101010101011 01\underline{0}1101101101010101101010101 \end{cases}$$

$$h_{31}^{+}\colon \begin{cases} a \to 1\underline{0}1010101101010101010101010101011\underline{0}110110110101010101010110110110 \\ b \to 1\underline{0}10101011010101010101010101010101101\underline{0}110110110101010101010110110101 \end{cases}$$

$$h_{32}^{+}\colon \begin{cases} a \to 1\underline{0}1110101011010110101101010101011\underline{0}11010101011010110110101010110 \\ b \to 1\underline{0}1110101011010110101101010101101\underline{0}11010101011010110110101010101 \end{cases}$$

$$h_{33}^{+}\colon \begin{cases} a \to 1\underline{0}10110101101101010101010101010101011\underline{0}1101011010110101010101010110 \\ b \to 1\underline{0}1011010110110101010101010101010101101\underline{0}1101011010110101010101010101 \end{cases}$$

$$h_{34}^{+}\colon \begin{cases} a \to 1\underline{0}10110101101010101010101010101010101011\underline{0}11010110101010101010101010110 \\ b \to 1\underline{0}1011010110101010101010101010101010101101\underline{0}11010110101010101010101010101 \end{cases}$$

$$h_{35}^{+}\colon \begin{cases} a \to 1\underline{0}1011011011010101010101010101010101011\underline{0}1101011011010101010101010101010110 \\ b \to 1\underline{0}101101101101010101010101010101010101101\underline{0}1101011011010101010101010101010101 \end{cases}$$

$$h_{36}^{+}\colon \begin{cases} a \to 1\underline{0}1011010101101010101010101010101010101 \\ \phantom{a \to{}} 1\underline{0}11010101011010101010101010101010110 \\ b \to 1\underline{0}1011010101101010101010101010101010110 \\ \phantom{b \to{}} 1\underline{0}11010101011010101010101010101010101 \end{cases}$$

$$h_{37}^{+}\colon \begin{cases} a \to 1\underline{0}1011010101010101010101010101010101 \\ \phantom{a \to{}} 1\underline{0}10101101010101010101010101010101010110 \\ b \to 1\underline{0}1011010101010101010101010101010101010110 \\ \phantom{b \to{}} 1\underline{0}10101101010101010101010101010101010101 \end{cases}$$

$$h_{38}^{+}\colon \begin{cases} a \to 1\underline{0}101010110101010110110101011010101010101 \\ \phantom{a \to{}} 1\underline{0}11011011010101010101010101011010101010110 \\ b \to 1\underline{0}101010110101010110110101011010101010110 \\ \phantom{b \to{}} 1\underline{0}11011011010101010101010101011010101010101 \end{cases}$$

Morphisms h_k^- for $9 \leq k \leq 38$

$$h_{9}^{-}\colon \begin{cases} a \to 1\underline{0}10110110\underline{1}... \end{cases}$$

$$h^-_{19}\colon \begin{cases} a \to 1\underline{0}1011011010110101101\underline{0}1101010110101101011\underline{0}101101101010110110 \\ b \to 1\underline{0}1110101011011010110\underline{1}1101010110101101101\underline{0}1011011010110101 01 \end{cases}$$

$$h^-_{20}\colon \begin{cases} a \to 1\underline{0}1011010110110110101\underline{1}0110101101101101011010\underline{1}011010101101010101 \\ b \to 1\underline{0}1101010110110110101\underline{1}0110101101101101010110\underline{1}011010101101010110 \end{cases}$$

$$h^-_{21}\colon \begin{cases} a \to 1\underline{0}101010101011010101101\underline{0}110110101010110101011\underline{0}11011010101011010110 \\ b \to 1\underline{0}110110101011010101101\underline{0}110110101010110101101\underline{0}10101010101101010101 \end{cases}$$

$$h^-_{22}\colon \begin{cases} a \to 1\underline{0}10101010110101011010110\underline{1}0110110110101011011010\underline{1}0101010101011011010101 \\ b \to 1\underline{0}1010101011010101101101\underline{0}101101101101010110101101\underline{0}101010101011011010110 \end{cases}$$

$$h^-_{23}\colon \begin{cases} a \to 1\underline{0}101010110101101101010110\underline{1}1011010110101010101010 \\ b \to 1\underline{0}1010101101011011010110\underline{1}01101101011010101010101 \end{cases}$$

$$h^-_{24}\colon \begin{cases} a \to 1\underline{0}101101011010101101101011\underline{0}110101101010110101100110 \\ b \to 1\underline{0}1011010110101101011011010\underline{1}1101011010101011010110101 \end{cases}$$

$$h^-_{25}\colon \begin{cases} a \to 1\underline{0}1010101101101011010101011\underline{0}1101101010101011010101 10 \\ b \to 1\underline{0}10101011011010110101011010\underline{1}101101010101011010101 01 \end{cases}$$

$$h^-_{26}\colon \begin{cases} a \to 1\underline{0}10101010101010101010101011\underline{0}1101101011011010101010110 \\ b \to 1\underline{0}1010101010101010101010110\underline{1}01101101011011010101010101 \end{cases}$$

$$h^-_{27}\colon \begin{cases} a \to 1\underline{0}1101010110101011010101011010\underline{1}10101101010101011010101 01 \\ b \to 1\underline{0}1101011010101011010101010110\underline{1}101010110101010110101 0110 \end{cases}$$

$$h^-_{28}\colon \begin{cases} a \to 1\underline{0}10110110110110101010101010110\underline{1}110101101101101010101010110 \\ b \to 1\underline{0}10110110110110101010101011010\underline{1}110101101101101010101010101 \end{cases}$$

$$h^-_{29}\colon \begin{cases} a \to 1\underline{0}1011010110110101010101010101011\underline{0}1010110110110101010101010110 \\ b \to 1\underline{0}10110101101101010101010101011010\underline{1}010110110110101010101010101 \end{cases}$$

$$h^-_{30}\colon \begin{cases} a \to 1\underline{0}10110101101011010110101010101011\underline{0}101011010110101101011010101 10 \\ b \to 1\underline{0}10110101101011010110101010101101\underline{0}101011010110101101011010101 01 \end{cases}$$

$$h^-_{31}\colon \begin{cases} a \to 1\underline{0}1011010101010101010101010101010101011\underline{0}11010101010101010101010101010110 \\ b \to 1\underline{0}101101010101010101010101010101101\underline{0}1101010101010101010101010101010101 \end{cases}$$

$$h^-_{32}\colon \begin{cases} a \to 1\underline{0}110110101101011010101010101010101011\underline{0}1101011010110101101010101010110 \\ b \to 1\underline{0}1101101011010110101010101010101011010\underline{1}101011010110101101010101010101 \end{cases}$$

$$h^-_{33}\colon \begin{cases} a \to 1\underline{0}101010110110110101010101010101010101011\underline{0}1101101101010101010101010101010110 \\ b \to 1\underline{0}101010110110110101010101010101010101101\underline{0}1101101101010101010101010101010101 \end{cases}$$

$$h^-_{34}\colon \begin{cases} a \to 1\underline{0}1011010110101101101010101010101 \\ 1\underline{0}10101101011011011010101010101010110 \\ b \to 1\underline{0}10110101101011011010101010101010110 \\ 1\underline{0}1010110101101101101010101010101010101 \end{cases}$$

$$h^-_{35}\colon \begin{cases} a \to 1\underline{0}10101011010110101101101101010101 \\ 1\underline{0}11011010110101101010101101010110 \\ b \to 1\underline{0}10101011010110101101101101010110 \\ 1\underline{0}11011010110101101010101101010101 \end{cases}$$

$$h^-_{36}\colon \begin{cases} a \to 1\underline{0}110110101101101010101010101010101 \\ 1\underline{0}1101011011011010101010101010101010110 \\ b \to 1\underline{0}110110101101101010101010101010110 \\ 1\underline{0}1101011011011010101010101010101010101 \end{cases}$$

$$h^-_{37}\colon \begin{cases} a \to 1\underline{0}1101101010110101010101010101010101 \\ 1\underline{0}11010110101101010101010101010101010110 \\ b \to 1\underline{0}1101101010110101010101010101010110 \\ 1\underline{0}11010110101101010101010101010101010101 \end{cases}$$

$$h^-_{38}\colon \begin{cases} a \to 1\underline{0}101010101010101101010101010101011010101 \\ 1\underline{0}101010101010101101010101010101011010110 \\ b \to 1\underline{0}101010101010101101010101010101011010110 \\ 1\underline{0}101010101010101101010101010101011010101 \end{cases}$$

Chapter 3

Ochem's conjecture

Since Dejean's conjecture is now proved, one can look at reinforcements of the question. This second chapter is dedicated to the proof of Ochem's conjecture, which adds an additional frequency constraint on a letter.

We use the method presented in the previous chapter. However, we cannot directly use the decision algorithm presented in Section 2.5, since the alphabet size is not fixed. The main part of this chapter is thus a proof that the presented constructions do not produce small repetitions (with respect to the alphabet size).

In all cases, our construction for the Pansiot code is a morphic word. Moreover, for alphabets with at least 9 letters, this morphic word is of the form $h(w_{TM})$, where w_{TM} is the Thue-Morse word, and h is a uniform morphism.

In addition to the proof of Ochem's conjecture, this result is the second known proof to Dejean's conjecture over large alphabet (the first one is from Carpi [49]).

This chapter is based on paper [32] (joint work with Elise Vaslet).

3.1 Preliminaries

Ochem introduced a stronger version of the conjecture involving letter frequencies. The *frequency* of a letter x in an infinite word w is $\lim_{n\to\infty} \frac{|w[1:n]|_x}{n}$, if the limit exists.

Conjecture 3.1 (Ochem's conjecture, [122]).

1. For every $k \geq 5$, there exists an infinite $\left(\frac{k}{k-1}^+\right)$-free word over a k-letter alphabet with letter frequency $\frac{1}{k+1}$.

2. For every $k \geq 6$, there exists an infinite $\left(\frac{k}{k-1}^+\right)$-free word over a k-letter alphabet with letter frequency $\frac{1}{k-1}$.

One can easily see that the frequencies of Ochem's conjecture are the best we can do. For other small alphabet sizes, the minimal and maximal frequencies are not known (see Section 5.1).

As we will see, the cases $9 \leq k \leq 38$ are proved by construction given in Chapter 2. We present in Section 3.2 a construction of words which prove Ochem's conjecture for every $k \geq 24$. As for Chapter 2, the Pansiot code of these words are image by a morphism of the Thue-Morse word. The Section 3.3 is devoted to the proof of result of Section 3.2 when $k \geq 32$. The cases when $k < 9$ are treated by specific constructions in Section 3.4. Thus, all cases of Ochem's conjecture are proved by results of the previous and this chapter.

We use the notations defined in Chapter 2. Nevertheless, in all this chapter, in order to simplify the following notations, the indices starts at 0.

Two words $u, v \in \Sigma^*$ are *isomorphic* if $|u| = |v|$ and there is a permutation $\sigma : \Sigma \to \Sigma$ such that $u[i] = \sigma(v[i])$ for every $i \in \{0, |u|-1\}$.

An *occurrence* of a factor u in w is an integer $i \in \mathbb{N}$ such that $w[i : i+|u|-1] = u$. The distance between two occurrences i and j is $|i-j|$. For example, the distance between the two occurrences of a in aba is 2.

Letter frequencies and Pansiot code. In all of the following, k denotes the size of the alphabet. We reuse the notations $P_k(w)$, $M_k(w)$ defined in Section 2.2.

Remark 3.2. If the Pansiot code of a Dejean word w has a 0 at every position $i \pmod{k-1}$ for an $i \in \{0, \dots k-1\}$, then w has the same letter at position $i \pmod{k-1}$, thus w has a letter with frequency $\frac{1}{k-1}$. Similarly if the Pansiot code of w has a 0 at every position $i \pmod{k+1}$ for an $i \in \{0, \dots k+1\}$, then w has a letter with frequency $\frac{1}{k+1}$.

By Remark 3.2, morphisms h_k^+ and h_k^- in Section 2.5 prove Ochem's conjecture for $k \in \{9, \dots 38\}$. The underlined 0 are those at position $i \pmod{k-1}$ (resp. $i \pmod{k+1}$).

We now introduce some notations in order to simplify forthcoming proofs.

Odd double crossed cycles. In this section, we fix a $k \in \mathbb{N}$.

The *invariants* of a permutation $\sigma \in \mathbb{S}_k$ are the $i \in \{0, \dots, k-1\}$ such that $\sigma(i) = i$. The set of invariants of σ is denoted $I(\sigma)$.

Definition 3.3 (Odd double crossed cycle)**.** A pair $(\sigma_a, \sigma_b) \in \mathbb{S}_k^2$ is a *double crossed cycle* if there is a $3 \le K \le \frac{k}{2}$ and two sequences $(a_0, \dots a_{K-1})$ and $(b_0, \dots, b_{K-1})$ of pairwise disjoint elements of Σ_k such that

$$\sigma_a = (a_0, a_1, a_2, \dots, a_{K-1})(b_0, b_1, b_2, \dots, b_{K-1})$$

and

$$\sigma_b = (b_0, a_1, a_2, \dots, a_{K-1})(a_0, b_1, b_2, \dots, b_{K-1}).$$

Moreover, (σ_a, σ_b) is an *odd double crossed cycle (ODCC)* if K is odd. The *size* of the double crossed cycle is K.

Lemma 3.4. *If (σ_a, σ_b) is an ODCC, then there is a $\sigma \in \mathbb{S}_k$ such that*

$$\sigma_a = \sigma\sigma_a\sigma_b\sigma^{-1}$$

and

$$\sigma_b = \sigma\sigma_b\sigma_a\sigma^{-1}.$$

Proof. Clearly, if $(\sigma_1, \sigma_2) \in \mathbb{S}_k^2$ and $(\sigma_3, \sigma_4) \in \mathbb{S}_k^2$ are two ODCC with the same size, then it exists a $\sigma \in \mathbb{S}_k$ such that $\sigma_1 = \sigma\sigma_3\sigma^{-1}$ and $\sigma_2 = \sigma\sigma_4\sigma^{-1}$. So it suffices to show that $(\sigma_a\sigma_b, \sigma_b\sigma_a)$ is an ODCC of size K. Let $\sigma_a = (a_0, a_1, \dots, a_{K-1})(b_0, b_1, \dots b_{K-1})$ and $\sigma_b = (b_0, a_1, \dots, a_{K-1})(a_0, b_1, \dots b_{K-1})$. Then

$$\sigma_b\sigma_a = (a_0, a_2, a_4 \dots a_{K-1}, b_1, b_3 \dots b_{K-2})(b_0, b_2, \dots b_{K-1}, a_1, a_3, \dots a_{K-2}),$$

and

$$\sigma_a\sigma_b = (b_0, a_2, a_4 \dots a_{K-1}, b_1, b_3 \dots b_{K-2})(a_0, b_2, \dots b_{K-1}, a_1, a_3, \dots a_{K-2}).$$

Thus $(\sigma_a\sigma_b, \sigma_b\sigma_a)$ is an ODCC of size K. □

Let (σ_a, σ_b) be an ODCC. The invariants of the ODCC are the invariants of σ_a and σ_b, and are denoted I when (σ_a, σ_b) is clear in the context. For every $i \in \{0, \ldots, K-1\}$, the *couple* i (or the *couple number* i) is (a_i, b_i), using notations of Definition 3.3.

Let $\varphi' : B^* \to \mathbb{S}_k$ be the morphism such that $\varphi'(0) = \sigma_a$, and $\varphi'(1) = \sigma_b$. The couple i is *good* in a binary word w if $\sigma(a_i) \in \{a_0, \ldots, a_{K-1}\}$, where $\sigma = \varphi'(w)$. Every couple is good in the empty word, and in 0^l for every l. Every couple is good, except 0 and $K-1$ in 1. Note that $\sigma = \mathrm{Id}_k$ if and only if $|w|$ is a multiple of K, and every couple is good in w.

On can easily compute good couples in w. Every 0 in w does not change the set of good couples, when every 1 changes the state of exactly two couples. Formally, we have:

Proposition 3.5. *Let w be a binary word of size multiple of K. The couple i is good in w if and only if the set $\{0 \le x < |w| : w[x] = 1$ and $x \in \{i-1, i\} \mod K\}$ has even cardinality.*

The *derivative word* of the binary word w, denoted w', is the binary word of size $|w|-1$ such that for every $i \in \{0, \ldots, |w'|-1\}$, $w'[i] = w[i] + w[i+1]$ (in GF(2)).

Proposition 3.6. *Let w be a binary word of size multiple of K. For $i \neq 0$, the couple i is good in w if and only if the set $\{0 \le x < |w|-1 : w'[x] = 1$ and $x \equiv i-1 \mod K\}$ has even cardinality.*

3.2 Main theorem and plan of the proof

Let $k \ge 24$ be an integer. Let $m \in \{0, 1, \ldots, 7\}$ be such that $k \equiv m \mod 8$. Let us define the uniform morphism $h_k^+ : \{a, b\}^* \to \{0, 1\}^*$ and $h_k^- : \{a, b\}^* \to \{0, 1\}^*$ as follows

$$h_k^+ : \begin{cases} 0 \mapsto & \iota_m (10)^{c^+} 101 \kappa_m (10)^{c^+} 110 \\ 1 \mapsto & \iota_m (10)^{c^+} 110 \kappa_m (10)^{c^+} 101, \end{cases}$$

$$h_k^- : \begin{cases} 0 \mapsto & \iota_{m-2} (10)^{c^-} 101 \kappa_{m-2} (10)^{c^-} 110 \\ 1 \mapsto & \iota_{m-2} (10)^{c^-} 110 \kappa_{m-2} (10)^{c^-} 101, \end{cases}$$

where ι_m and κ_m are finite words over alphabet $\{0, 1\}$ defined in the Table 3.1 (the indices are taken modulo 8), $c^+ = \frac{1}{4}(2k - 4 - |\iota_m| - |\kappa_m|)$, and $c^- = \frac{1}{4}(2k - 8 - |\iota_{m-2}| - |\kappa_{m-2}|)$.

Theorem 3.7. *For any integer $k \ge 24$, the morphism h_k^+ (resp. h_k^-) is such that the infinite word $w_k^+ = M_k(h_k^+(w_{TM}))$ (resp. $w_k^- = M_k(h_k^-(w_{TM}))$) over Σ_k is $RT(k)^+$-free and have a letter of frequency $\frac{1}{k+1}$ (resp. $\frac{1}{k-1}$).*

By Remark 3.2, we know that $M_k(h_k^+(w_{TM}))$ has a letter of frequency $\frac{1}{k+1}$, since $h_k^+(0)[1] = h_k^+(1)[1] = h_k^+(0)[1+k+1] = h_k^+(1)[1+k+1] = 0$, and $|h_k^+(0)| = |h_k^+(1)| = k+1$. Similarly, we know that $M_k(h_k^-(w_{TM}))$ has a letter of frequency $\frac{1}{k-1}$. Thus, to prove Theorem 3.7, we only have to prove that $M_k(h_k^+(w_{TM}))$ and $M_k(h_k^-(w_{TM}))$ are Dejean words. This fact is proved by

m	ι_m	κ_m
0	101101010110101101	101011011010110101
1	101101010110101	101011011010110
2	1011010101101101	1010110110101101
3	1011010101101	1010110110101
4	1011010101	1010110110
5	10110101101	10101101101
6	10110101	10101101
7	10110	10101

Table 3.1 – Construction for the infinite case.

computer for every $24 \leq k \leq 31$, using the method presented in Chapter 2. We present in the following a proof for every $k \geq 32$.

To prove that w (where w is either $M_k(h_k^-(w_{TM}))$ or $M_k(h_k^+(w_{TM}))$ for a $k \geq 32$) is $RT(k)^+$-free, we will use a method similar to Pansiot's one, by considering different type of repetitions and dealing with each type separately. We give here the structure of the proof, which will be completed in next section.

- We show that w has no forbidden short repetition of excess less than 6, using characterizations of Carpi, Shur and Gorbunova.
- If w has a forbidden repetition of excess at least 5, then the period is a multiple of $s{\cdot}K$ (Section 3.3.1), where K is the size of the ODCC $(\varphi(h_k^+(0)), \varphi(h_k^+(1)))$ or $(\varphi(h_k^-(0)), \varphi(h_k^-(1)))$.
- w has no forbidden repetition of period $s{\cdot}K$ (Section 3.3.1).
- A forbidden repetition of period at least $2{\cdot}s{\cdot}K$ is a kernel repetition. Using properties of the derivative of the Thue-Morse word, we show that a forbidden kernel repetition has period at least $18{\cdot}s{\cdot}K$ (Section 3.3.2).
- If w has a forbidden kernel repetition of period $|p| \geq 18{\cdot}s{\cdot}K$, then w has a forbidden repetition of period $|p|/2$ (Section 3.3.3). This part follows the proof in Chapter 2.

3.3 Proof for large alphabets

We fix $k \geq 32$ and $\Delta \in \{-1, 1\}$. Let $\iota_a = \iota_m(10)^c101$, $\iota_a = \iota_m(10)^c110$, $\kappa_a = \kappa_m(10)^c101$ and $\kappa_a = \kappa_m(10)^c110$, where $c = \frac{1}{4}(2(k+\Delta) - 6 - |\iota_m| - |\kappa_m|)$ and $k + (\Delta - 1) \equiv m \mod 8$. Let $\lambda_a = \iota_a\kappa_b$ and $\lambda_b = \iota_b\kappa_a$. Let $h : B^* \to \Sigma_k^*$ be the morphism such that $h(0) = \lambda_a$ and $h(1) = \lambda_b$. Let $s = |h(0)| = |h(1)| = 2(k+\Delta)$. The largest common prefix L of $h(0)$ and $h(1)$ has size $\ell = \frac{s}{2} - 2$, and their last letter differ.

Note that $h = h_k^+$ if $\Delta = 1$ and $h = h_k^-$ if $\Delta = -1$. We prove that $M_k(h(w_{TM}))$ is a Dejean word,

Let $\sigma_a = \varphi(h(0))$, $\sigma_b = \varphi(h(1))$ and $\varphi' : B^* \to \mathbb{S}_k$ be the morphism such that $\varphi'(0) = \sigma_a$, and $\varphi'(1) = \sigma_b$. The permutations σ_a and σ_b are given in Table 3.2

for every k modulo 8 and Δ. Each permutations can be easily computed by hand using the definition of φ, or by a computer program. We see directly:

Fact 3.8. (σ_a, σ_b) is an ODCC of size at least $\lceil \frac{k}{4} \rceil$.

From now on, K denotes the size of the ODCC (σ_a, σ_b), I its invariants, and a_i and b_i $(0 \leq i < K)$ denote elements of cycles of (σ_a, σ_b) as in Definition 3.3. Table 3.2 also present the size of the ODCC and its invariants.

3.3.1 Avoiding short repetitions

In this section, we suppose that $\mathbf{w}' = h(\mathbf{u})$ for an infinite binary word $\mathbf{u}$, and that $\mathbf{w} = M_k(\mathbf{w}')$. Moreover, we suppose that $\mathbf{u}$ is recurrent.

Lemma 3.9. $\mathbf{w}$ *has no forbidden repetition of excess at most* 5*.*

Proof. We use characterization of stabilizing words of Carpi, Shur and Gorbunova [49, 140]. Since $\mathbf{w}'$ is $(00, 111)$-free, $\mathbf{w}$ has no forbidden repetition of excess 1 and 2. $\mathbf{w}'$ does not contains a factor of the form $uB^{k-4}u$, for a $u \in B^4$. Thus by Carpi's characterization of 3-stabilizing words, $\mathbf{w}$ has no forbidden repetition of excess 3 [49]. Finally, Shur and Gorbunova showed that a word on alphabet $k \geq 8$ which has no forbidden repetitions of excess 1, 2 and 3 has no forbidden repetition of excess 4 and 5 [140]. □

Let $u_a = (M_k(h(0)))[0 : s-1]$ and $u_b = (M_k(h(1)))[0 : s-1]$. Let $\mathcal{E} = \{i \in \{0, \ldots, s-1\} : u_a[i] = u_b[i]\}$. One can show on Table 3.2 that $\{0, \ldots, k-5\} \subseteq \mathcal{E}$. The word $\mathbf{w}$ is a concatenation of words isomorphic to u_a and u_b. For every n, $\mathbf{w}[n{\cdot}s : n{\cdot}s+k-2]$ is uniquely determined by $\varphi(h(\mathbf{u}[0 : n-1])) = \varphi'(\mathbf{u}[0 : n-1])$. More precisely:

Proposition 3.10. *For every* $n \geq 0$ *and* $i \in \{0, \ldots, k-2\}$, $\mathbf{w}[n{\cdot}s+i] = \varphi'(\mathbf{u}[0 : n-1])[i]$*.*

For every $i \in I$ and $n \geq 0$, we have $\mathbf{w}[i + n{\cdot}s] = i$ (note that $k-1 \notin I$). Since $s = 2k+2$ or $s = 2k-2$, there is an other occurrence of i exactly in mid-position, and we have $\mathbf{w}[i + n{\cdot}s + \frac{s}{2}] = i$. Thus every letter at position $\mathcal{I} = I + \frac{s}{2}{\cdot}\mathbb{N}$ is independent of $\mathbf{u}$.

Let $\mathcal{J} = (\mathcal{E} \setminus \mathcal{I}) + s{\cdot}\mathbb{N}$, and let $\gamma : \mathcal{J} \to \{0, \ldots K-1\}$ such that $\mathbf{w}[j] \in \{a_{\gamma(j)}, b_{\gamma(j)}\}$. Note that γ is well defined, and is independent of $\mathbf{u}$, by the properties of ODCC. By definition of φ', one has the following.

Proposition 3.11. $\gamma(x+s) = 1 + \gamma(x)$ *(modulo K).*

Looking at the permutations in Table 3.2, we see:

Fact 3.12. For every $x \in \mathbb{N}$, $\{x, \ldots, x+4\} \cap \mathcal{I} \neq \emptyset$ and $\{x, \ldots, x+4\} \cap \mathcal{J} \neq \emptyset$.

Suppose that $\mathbf{w}$ has a forbidden repetition. Let l be the least integer such that $\mathbf{w}$ has a forbidden repetition of period l. Let (p, e) be an forbidden repetition of period l, and let t be an integer such that $\mathbf{w}[t : \infty]$ has pe as prefix. Suppose w.l.o.g than $t \geq s{\cdot}K$ (since $\mathbf{u}$ is recurrent, if $\mathbf{w}$ has a forbidden repetition of period l, it has infinitely many forbidden repetitions of period l).

Lemma 3.13. *Let* $i \in I$ *and* $j \in \{0, \ldots k-1\} \setminus I$*.*

Case $k \equiv 0 \mod 8$, $\Delta = 1$:

$$\begin{array}{rcl}
\sigma_a & = & (k{-}2\ \ 1\ \ 3\ \ 6\ \ 7\ \ 11\ \ 12\ \ 16\ \ 19\ \ ..(4)..\ \ k{-}9\ \ k{-}5) \\
& & (k{-}1\ \ 2\ \ 4\ \ 8\ \ 9\ \ 13\ \ 14\ \ 17\ \ 21\ \ ..(4)..\ \ k{-}7\ \ k{-}3) \\
\sigma_b & = & (k{-}1\ \ 1\ \ 3\ \ 6\ \ 7\ \ 11\ \ 12\ \ 16\ \ 19\ \ ..(4)..\ \ k{-}9\ \ k{-}5) \\
& & (k{-}2\ \ 2\ \ 4\ \ 8\ \ 9\ \ 13\ \ 14\ \ 17\ \ 21\ \ ..(4)..\ \ k{-}7\ \ k{-}3) \\
I & = & \{0,5,10,15,18,..(2)..,k{-}6,k{-}4\} \quad K = \frac{k+12}{4}
\end{array}$$

1 2 3 4 5 6 7 8 -2 -1 0 1

$$u_a = \mathbf{a}\,b\,c\,d\,e\,\mathbf{f}\,g\,h\,i\,j\,\mathbf{k}\,l\,m\,n\,o\,\mathbf{p}\,q\,r\,\mathbf{s}\,t\,\mathbf{u}\,v\ \dots\ \underline{\mathbf{q}}\,\underline{r}\,\underline{\mathbf{s}}\,\underline{t}\,\underline{\mathbf{u}}\,\underline{v}\,\underline{\mathbf{w}}\,\underline{x}\,\underline{y}\,\underline{z}\,b$$

1 2 3 4 5 6 7 8 9 -2 -1 0 1

$$\mathbf{a}\,c\,e\,d\,g\,\mathbf{f}\,i\,h\,j\,l\,\mathbf{k}\,n\,m\,o\,q\,\mathbf{p}\,r\,t\,\mathbf{s}\,v\,\mathbf{u}\,x\ \dots\ \underline{\mathbf{q}}\,\underline{t}\,\underline{\mathbf{s}}\,\underline{v}\,\underline{\mathbf{u}}\,\underline{x}\,\underline{\mathbf{w}}\,\underline{y}\,b\,\underline{z}\,c$$

1 2 3 4 5 6 7 8 -2 -1 0 1

$$u_b = \mathbf{a}\,b\,c\,d\,e\,\mathbf{f}\,g\,h\,i\,j\,\mathbf{k}\,l\,m\,n\,o\,\mathbf{p}\,q\,r\,\mathbf{s}\,t\,\mathbf{u}\,v\ \dots\ \underline{\mathbf{q}}\,\underline{r}\,\underline{\mathbf{s}}\,\underline{t}\,\underline{\mathbf{u}}\,\underline{v}\,\underline{\mathbf{w}}\,\underline{x}\,\underline{y}\,\underline{z}\,b$$

1 2 3 4 5 6 7 8 9 -2 -1 0 1

$$\mathbf{a}\,c\,e\,d\,g\,\mathbf{f}\,i\,h\,j\,l\,\mathbf{k}\,n\,m\,o\,q\,\mathbf{p}\,r\,t\,\mathbf{s}\,v\,\mathbf{u}\,x\ \dots\ \underline{\mathbf{q}}\,\underline{t}\,\underline{\mathbf{s}}\,\underline{v}\,\underline{\mathbf{u}}\,\underline{x}\,\underline{\mathbf{w}}\,\underline{z}\,\underline{y}\,b\,c$$

Case $k \equiv 1 \mod 8$, $\Delta = 1$:

$$\begin{array}{rcl}
\sigma_a & = & (k{-}2\ \ 1\ \ 3\ \ 6\ \ 7\ \ 11\ \ 12\ \ 16\ \ ..(4)..\ \ k{-}9\ \ k{-}5) \\
& & (k{-}1\ \ 2\ \ 4\ \ 8\ \ 9\ \ 13\ \ 14\ \ 18\ \ ..(4)..\ \ k{-}7\ \ k{-}3) \\
\sigma_b & = & (k{-}1\ \ 1\ \ 3\ \ 6\ \ 7\ \ 11\ \ 12\ \ 16\ \ ..(4)..\ \ k{-}9\ \ k{-}5) \\
& & (k{-}2\ \ 2\ \ 4\ \ 8\ \ 9\ \ 13\ \ 14\ \ 18\ \ ..(4)..\ \ k{-}7\ \ k{-}3) \\
I & = & \{0,5,10,15,..(2)..,k{-}6,k{-}4\} \quad K = \frac{k+11}{4}
\end{array}$$

1 2 3 4 5 6 7 -2 -1 0 1

$$u_a = \mathbf{a}\,b\,c\,d\,e\,\mathbf{f}\,g\,h\,i\,j\,\mathbf{k}\,l\,m\,n\,o\,\mathbf{p}\,q\,\mathbf{r}\,s\ \dots\ \underline{\mathbf{q}}\,\underline{r}\,\underline{\mathbf{s}}\,\underline{t}\,\underline{\mathbf{u}}\,\underline{v}\,\underline{\mathbf{w}}\,\underline{x}\,\underline{y}\,\underline{z}\,b$$

1 2 3 4 5 6 7 8 -2 -1 0 1

$$\mathbf{a}\,c\,e\,d\,g\,\mathbf{f}\,i\,h\,j\,l\,\mathbf{k}\,n\,m\,o\,q\,\mathbf{p}\,s\,\mathbf{r}\,u\ \dots\ \underline{\mathbf{q}}\,\underline{t}\,\underline{\mathbf{s}}\,\underline{v}\,\underline{\mathbf{u}}\,\underline{x}\,\underline{\mathbf{w}}\,\underline{y}\,b\,\underline{z}\,c$$

1 2 3 4 5 6 7 -2 -1 0 1

$$u_b = \mathbf{a}\,b\,c\,d\,e\,\mathbf{f}\,g\,h\,i\,j\,\mathbf{k}\,l\,m\,n\,o\,\mathbf{p}\,q\,\mathbf{r}\,s\ \dots\ \underline{\mathbf{q}}\,\underline{r}\,\underline{\mathbf{s}}\,\underline{t}\,\underline{\mathbf{u}}\,\underline{v}\,\underline{\mathbf{w}}\,\underline{x}\,\underline{y}\,\underline{z}\,b$$

1 2 3 4 5 6 7 8 -2 -1 0 1

$$\mathbf{a}\,c\,e\,d\,g\,\mathbf{f}\,i\,h\,j\,l\,\mathbf{k}\,n\,m\,o\,q\,\mathbf{p}\,s\,\mathbf{r}\,u\ \dots\ \underline{\mathbf{q}}\,\underline{t}\,\underline{\mathbf{s}}\,\underline{v}\,\underline{\mathbf{u}}\,\underline{x}\,\underline{\mathbf{w}}\,\underline{z}\,\underline{y}\,b\,c$$

Case $k \equiv 2 \mod 8$, $\Delta = 1$:

$$\begin{array}{rcl}
\sigma_a & = & (k{-}2\ \ 1\ \ 3\ \ 6\ \ 7\ \ 11\ \ 14\ \ 17\ \ ..(4)..\ \ k{-}9\ \ k{-}5) \\
& & (k{-}1\ \ 2\ \ 4\ \ 8\ \ 9\ \ 12\ \ 15\ \ 19\ \ ..(4)..\ \ k{-}7\ \ k{-}3) \\
\sigma_b & = & (k{-}1\ \ 1\ \ 3\ \ 6\ \ 7\ \ 11\ \ 14\ \ 17\ \ ..(4)..\ \ k{-}9\ \ k{-}5) \\
& & (k{-}2\ \ 2\ \ 4\ \ 8\ \ 9\ \ 12\ \ 15\ \ 19\ \ ..(4)..\ \ k{-}7\ \ k{-}3) \\
I & = & \{0,5,10,13,16,..(2)..,k{-}6,k{-}4\} \quad K = \frac{k+10}{4}
\end{array}$$

1 2 3 4 5 6 7 -2 -1 0 1

$$u_a = \mathbf{a}\,b\,c\,d\,e\,\mathbf{f}\,g\,h\,i\,j\,\mathbf{k}\,l\,m\,\mathbf{n}\,o\,p\,\mathbf{q}\,r\,\mathbf{s}\,t\ \dots\ \underline{\mathbf{q}}\,\underline{r}\,\underline{\mathbf{s}}\,\underline{t}\,\underline{\mathbf{u}}\,\underline{v}\,\underline{\mathbf{w}}\,\underline{x}\,\underline{y}\,\underline{z}\,b$$

1 2 3 4 5 6 7 8 -2 -1 0 1

$$\mathbf{a}\,c\,e\,d\,g\,\mathbf{f}\,i\,h\,j\,l\,\mathbf{k}\,m\,o\,\mathbf{n}\,p\,r\,\mathbf{q}\,t\,\mathbf{s}\,v\ \dots\ \underline{\mathbf{q}}\,\underline{t}\,\underline{\mathbf{s}}\,\underline{v}\,\underline{\mathbf{u}}\,\underline{x}\,\underline{\mathbf{w}}\,\underline{y}\,b\,\underline{z}\,c$$

1 2 3 4 5 6 7 -2 -1 0 1

$$u_b = \mathbf{a}\,b\,c\,d\,e\,\mathbf{f}\,g\,h\,i\,j\,\mathbf{k}\,l\,m\,\mathbf{n}\,o\,p\,\mathbf{q}\,r\,\mathbf{s}\,t\ \dots\ \underline{\mathbf{q}}\,\underline{r}\,\underline{\mathbf{s}}\,\underline{t}\,\underline{\mathbf{u}}\,\underline{v}\,\underline{\mathbf{w}}\,\underline{x}\,\underline{y}\,\underline{z}\,b$$

1 2 3 4 5 6 7 8 -2 -1 0 1

$$\mathbf{a}\,c\,e\,d\,g\,\mathbf{f}\,i\,h\,j\,l\,\mathbf{k}\,m\,o\,\mathbf{n}\,p\,r\,\mathbf{q}\,t\,\mathbf{s}\,v\ \dots\ \underline{\mathbf{q}}\,\underline{t}\,\underline{\mathbf{s}}\,\underline{v}\,\underline{\mathbf{u}}\,\underline{x}\,\underline{\mathbf{w}}\,\underline{z}\,\underline{y}\,b\,c$$

Case $k \equiv 3 \mod 8$, $\Delta = 1$:

$$\begin{array}{rcl}
\sigma_a & = & (k{-}2\quad 1\quad 3\quad 6\quad 7\quad 11\quad 14\quad ..(4)..\quad k{-}9\quad k{-}5) \\
 & & (k{-}1\quad 2\quad 4\quad 8\quad 9\quad 12\quad 16\quad ..(4)..\quad k{-}7\quad k{-}3) \\
\sigma_b & = & (k{-}1\quad 1\quad 3\quad 6\quad 7\quad 11\quad 14\quad ..(4)..\quad k{-}9\quad k{-}5) \\
 & & (k{-}2\quad 2\quad 4\quad 8\quad 9\quad 12\quad 16\quad ..(4)..\quad k{-}7\quad k{-}3) \\
I & = & \{0, 5, 10, 13, ..(2).., k{-}6, k{-}4\} \quad K = \frac{k+9}{4}
\end{array}$$

1 2 3 4 5 6 ... -2 -1 0 1

$u_a = \mathbf{a}\,b\,c\,d\,e\,\mathbf{f}\,g\,h\,i\,j\,\mathbf{k}\,l\,m\,\mathbf{n}\,o\,\mathbf{p}\,q \ldots \underline{\mathbf{q}}\,\underline{r}\,\underline{\mathbf{s}}\,\underline{t}\,\underline{\mathbf{u}}\,\underline{v\mathbf{w}x}\,\underline{y}\,\underline{z}\,b$

1 2 3 4 5 6 7 ... -2 -1 0 1

$\quad\;\; \mathbf{a}\,c\,e\,d\,g\,\mathbf{f}\,i\,h\,j\,l\,\mathbf{k}\,m\,o\,\mathbf{n}\,q\,\mathbf{p}\,s \ldots \underline{\mathbf{q}}\,\underline{t}\,\underline{s}\,\underline{v}\,\underline{\mathbf{u}}\,\underline{x\mathbf{w}y}\,b\,\underline{z}\,c$

1 2 3 4 5 6 ... -2 -1 0 1

$u_b = \mathbf{a}\,b\,c\,d\,e\,\mathbf{f}\,g\,h\,i\,j\,\mathbf{k}\,l\,m\,\mathbf{n}\,o\,\mathbf{p}\,q \ldots \underline{\mathbf{q}}\,\underline{r}\,\underline{\mathbf{s}}\,\underline{t}\,\underline{\mathbf{u}}\,\underline{v\mathbf{w}x}\,\underline{y}\,\underline{z}\,b$

1 2 3 4 5 6 7 ... -2 -1 0 1

$\quad\;\; \mathbf{a}\,c\,e\,d\,g\,\mathbf{f}\,i\,h\,j\,l\,\mathbf{k}\,m\,o\,\mathbf{n}\,q\,\mathbf{p}\,s \ldots \underline{\mathbf{q}}\,\underline{t}\,\underline{s}\,\underline{v}\,\underline{\mathbf{u}}\,\underline{x\mathbf{w}z}\,\underline{y}\,b\,c$

Case $k \equiv 4 \mod 8$, $\Delta = 1$:

$$\begin{array}{rcl}
\sigma_a & = & (k{-}2\quad 1\quad 3\quad 6\quad 7\quad 11\quad ..(4)..\quad k{-}9\quad k{-}5) \\
 & & (k{-}1\quad 2\quad 4\quad 8\quad 9\quad 13\quad ..(4)..\quad k{-}7\quad k{-}3) \\
\sigma_b & = & (k{-}1\quad 1\quad 3\quad 6\quad 7\quad 11\quad ..(4)..\quad k{-}9\quad k{-}5) \\
 & & (k{-}2\quad 2\quad 4\quad 8\quad 9\quad 13\quad ..(4)..\quad k{-}7\quad k{-}3) \\
I & = & \{0, 5, 10, ..(2).., k{-}6, k{-}4\} \quad K = \frac{k+8}{4}
\end{array}$$

1 2 3 4 5 ... -2 -1 0 1

$u_a = \mathbf{a}\,b\,c\,d\,e\,\mathbf{f}\,g\,h\,i\,j\,\mathbf{k}\,l\,\mathbf{m}\,n \ldots \underline{\mathbf{q}}\,\underline{r}\,\underline{\mathbf{s}}\,\underline{t}\,\underline{\mathbf{u}}\,\underline{v\mathbf{w}x}\,\underline{y}\,\underline{z}\,b$

1 2 3 4 5 6 ... -2 -1 0 1

$\quad\;\; \mathbf{a}\,c\,e\,d\,g\,\mathbf{f}\,i\,h\,j\,l\,\mathbf{k}\,n\,\mathbf{m}\,p \ldots \underline{\mathbf{q}}\,\underline{t}\,\underline{s}\,\underline{v}\,\underline{\mathbf{u}}\,\underline{x\mathbf{w}y}\,b\,\underline{z}\,c$

1 2 3 4 5 ... -2 -1 0 1

$u_b = \mathbf{a}\,b\,c\,d\,e\,\mathbf{f}\,g\,h\,i\,j\,\mathbf{k}\,l\,\mathbf{m}\,n \ldots \underline{\mathbf{q}}\,\underline{r}\,\underline{\mathbf{s}}\,\underline{t}\,\underline{\mathbf{u}}\,\underline{v\mathbf{w}x}\,\underline{y}\,\underline{z}\,b$

1 2 3 4 5 6 ... -2 -1 0 1

$\quad\;\; \mathbf{a}\,c\,e\,d\,g\,\mathbf{f}\,i\,h\,j\,l\,\mathbf{k}\,n\,\mathbf{m}\,p \ldots \underline{\mathbf{q}}\,\underline{t}\,\underline{s}\,\underline{v}\,\underline{\mathbf{u}}\,\underline{x\mathbf{w}z}\,\underline{y}\,b\,c$

Case $k \equiv 5 \mod 8$, $\Delta = 1$:

$$\begin{array}{rcl}
\sigma_a & = & (k{-}2\quad 1\quad 3\quad 6\quad 9\quad 12\quad ..(4)..\quad k{-}9\quad k{-}5) \\
 & & (k{-}1\quad 2\quad 4\quad 7\quad 10\quad 14\quad ..(4)..\quad k{-}7\quad k{-}3) \\
\sigma_b & = & (k{-}1\quad 1\quad 3\quad 6\quad 9\quad 12\quad ..(4)..\quad k{-}9\quad k{-}5) \\
 & & (k{-}2\quad 2\quad 4\quad 7\quad 10\quad 14\quad ..(4)..\quad k{-}7\quad k{-}3) \\
I & = & \{0, 5, 8, 11, ..(2).., k{-}6, k{-}4\} \quad K = \frac{k+7}{4}
\end{array}$$

1 2 3 4 5 ... -2 -1 0 1

$u_a = \mathbf{a}\,b\,c\,d\,e\,\mathbf{f}\,g\,h\,\mathbf{i}\,j\,k\,\mathbf{l}\,m\,\mathbf{n}\,o \ldots \underline{\mathbf{q}}\,\underline{r}\,\underline{\mathbf{s}}\,\underline{t}\,\underline{\mathbf{u}}\,\underline{v\mathbf{w}x}\,\underline{y}\,\underline{z}\,b$

1 2 3 4 5 6 ... -2 -1 0 1

$\quad\;\; \mathbf{a}\,c\,e\,d\,g\,\mathbf{f}\,h\,j\,\mathbf{i}\,k\,m\,\mathbf{l}\,o\,\mathbf{n}\,q \ldots \underline{\mathbf{q}}\,\underline{t}\,\underline{s}\,\underline{v}\,\underline{\mathbf{u}}\,\underline{x\mathbf{w}y}\,b\,\underline{z}\,c$

1 2 3 4 5 ... -2 -1 0 1

$u_b = \mathbf{a}\,b\,c\,d\,e\,\mathbf{f}\,g\,h\,\mathbf{i}\,j\,k\,\mathbf{l}\,m\,\mathbf{n}\,o \ldots \underline{\mathbf{q}}\,\underline{r}\,\underline{\mathbf{s}}\,\underline{t}\,\underline{\mathbf{u}}\,\underline{v\mathbf{w}x}\,\underline{y}\,\underline{z}\,b$

1 2 3 4 5 6 ... -2 -1 0 1

$\quad\;\; \mathbf{a}\,c\,e\,d\,g\,\mathbf{f}\,h\,j\,\mathbf{i}\,k\,m\,\mathbf{l}\,o\,\mathbf{n}\,q \ldots \underline{\mathbf{q}}\,\underline{t}\,\underline{s}\,\underline{v}\,\underline{\mathbf{u}}\,\underline{x\mathbf{w}z}\,\underline{y}\,b\,c$

Case $k \equiv 6 \mod 8$, $\Delta = 1$:

$$\begin{array}{rcl}
\sigma_a & = & (k{-}2 \quad 1 \quad 3 \quad 6 \quad 9 \quad ..(4).. \quad k{-}9 \quad k{-}5) \\
 & & (k{-}1 \quad 2 \quad 4 \quad 7 \quad 11 \quad ..(4).. \quad k{-}7 \quad k{-}3) \\
\sigma_b & = & (k{-}1 \quad 1 \quad 3 \quad 6 \quad 9 \quad ..(4).. \quad k{-}9 \quad k{-}5) \\
 & & (k{-}2 \quad 2 \quad 4 \quad 7 \quad 11 \quad ..(4).. \quad k{-}7 \quad k{-}3) \\
I & = & \{0, 5, 8, ..(2).., k{-}6, k{-}4\} \quad K = \frac{k+6}{4}
\end{array}$$

$$\begin{array}{rcl}
 & & \quad 1 \quad 2 \quad\ 3 \quad\ \ 4 \qquad\qquad\quad -2 \quad -1 \quad 0 \quad 1 \\
u_a & = & \mathbf{a}\,b\,c\,d\,e\,\mathbf{f}\,g\,h\,\mathbf{i}\,j\,\mathbf{k}\,l\ \ldots\ \underline{\mathbf{q}}\,\underline{r}\,\underline{\mathbf{s}}\,\underline{t}\,\underline{\mathbf{u}}\,\underline{v\mathbf{w}x}\,\underline{y}\,\underline{z}\,b \\
 & & 1 \quad 2 \quad 3 \quad\ 4 \quad\ 5 \qquad\quad -2 \quad -1 \quad\ 0 \quad 1 \\
 & & \mathbf{a}\,c\,e\,d\,g\,\mathbf{f}\,h\,j\,\mathbf{i}\,l\,\mathbf{k}\,n\ \ldots\ \underline{\mathbf{q}}\,\underline{t}\,\underline{\mathbf{s}}\,\underline{v}\,\underline{\mathbf{u}}\,\underline{x\mathbf{w}}\,\underline{y}\,b\,\underline{z}\,c \\
 & & \\
 & & \quad 1 \quad 2 \quad\ 3 \quad\ \ 4 \qquad\qquad\quad -2 \quad -1 \quad 0 \quad 1 \\
u_b & = & \mathbf{a}\,b\,c\,d\,e\,\mathbf{f}\,g\,h\,\mathbf{i}\,j\,\mathbf{k}\,l\ \ldots\ \underline{\mathbf{q}}\,\underline{r}\,\underline{\mathbf{s}}\,\underline{t}\,\underline{\mathbf{u}}\,\underline{v\mathbf{w}x}\,\underline{y}\,\underline{z}\,b \\
 & & 1 \quad 2 \quad 3 \quad\ 4 \quad\ 5 \qquad\quad -2 \quad -1 \quad\ 0 \quad 1 \\
 & & \mathbf{a}\,c\,e\,d\,g\,\mathbf{f}\,h\,j\,\mathbf{i}\,l\,\mathbf{k}\,n\ \ldots\ \underline{\mathbf{q}}\,\underline{t}\,\underline{\mathbf{s}}\,\underline{v}\,\underline{\mathbf{u}}\,\underline{x\mathbf{w}z}\,\underline{y}\,b\,c
\end{array}$$

Case $k \equiv 7 \mod 8$, $\Delta = 1$:

$$\begin{array}{rcl}
\sigma_a & = & (k{-}2 \quad 1 \quad 3 \quad 6 \quad ..(4).. \quad k{-}9 \quad k{-}5) \\
 & & (k{-}1 \quad 2 \quad 4 \quad 8 \quad ..(4).. \quad k{-}7 \quad k{-}3) \\
\sigma_b & = & (k{-}1 \quad 1 \quad 3 \quad 6 \quad ..(4).. \quad k{-}9 \quad k{-}5) \\
 & & (k{-}2 \quad 2 \quad 4 \quad 8 \quad ..(4).. \quad k{-}7 \quad k{-}3) \\
I & = & \{0, 5, ..(2).., k{-}6, k{-}4\} \quad K = \frac{k+5}{4}
\end{array}$$

$$\begin{array}{rcl}
 & & \quad 1 \quad 2 \quad\ 3 \qquad\qquad -2 \quad -1 \quad 0 \quad 1 \\
u_a & = & \mathbf{a}\,b\,c\,d\,e\,\mathbf{f}\,g\,\mathbf{h}\,i\ \ldots\ \underline{\mathbf{q}}\,\underline{r}\,\underline{\mathbf{s}}\,\underline{t}\,\underline{\mathbf{u}}\,\underline{v\mathbf{w}x}\,\underline{y}\,\underline{z}\,b \\
 & & 1 \quad 2 \quad 3 \quad\ 4 \qquad -2 \quad -1 \quad\ 0 \quad 1 \\
 & & \mathbf{a}\,c\,e\,d\,g\,\mathbf{f}\,i\,\mathbf{h}\,k\ \ldots\ \underline{\mathbf{q}}\,\underline{t}\,\underline{\mathbf{s}}\,\underline{v}\,\underline{\mathbf{u}}\,\underline{x\mathbf{w}}\,\underline{y}\,b\,\underline{z}\,c \\
 & & \\
 & & \quad 1 \quad 2 \quad\ 3 \qquad\qquad -2 \quad -1 \quad 0 \quad 1 \\
u_b & = & \mathbf{a}\,b\,c\,d\,e\,\mathbf{f}\,g\,\mathbf{h}\,i\ \ldots\ \underline{\mathbf{q}}\,\underline{r}\,\underline{\mathbf{s}}\,\underline{t}\,\underline{\mathbf{u}}\,\underline{v\mathbf{w}x}\,\underline{y}\,\underline{z}\,b \\
 & & 1 \quad 2 \quad 3 \quad\ 4 \qquad -2 \quad -1 \quad\ 0 \quad 1 \\
 & & \mathbf{a}\,c\,e\,d\,g\,\mathbf{f}\,i\,\mathbf{h}\,k\ \ldots\ \underline{\mathbf{q}}\,\underline{t}\,\underline{\mathbf{s}}\,\underline{v}\,\underline{\mathbf{u}}\,\underline{x\mathbf{w}z}\,\underline{y}\,b\,c
\end{array}$$

Case $k \equiv 0 \mod 8$, $\Delta = -1$:

$$\begin{array}{rcl}
\sigma_a &=& (k{-}3 \quad k{-}4 \quad ..(-4).. \quad 24 \quad 20 \quad 16 \quad 13 \quad 10 \quad 8 \quad 5 \quad 4 \quad 0) \\
 && (k{-}2 \quad k{-}6 \quad ..(-4).. \quad 22 \quad 18 \quad 15 \quad 12 \quad 9 \quad 7 \quad 3 \quad 2 \quad k{-}1) \\
\sigma_b &=& (k{-}2 \quad k{-}4 \quad ..(-4).. \quad 24 \quad 20 \quad 16 \quad 13 \quad 10 \quad 8 \quad 5 \quad 4 \quad 0) \\
 && (k{-}3 \quad k{-}6 \quad ..(-4).. \quad 22 \quad 18 \quad 15 \quad 12 \quad 9 \quad 7 \quad 3 \quad 2 \quad k{-}1) \\
I &=& \{1,6,11,14,17,19,..(2)..,k{-}5\} \quad K = \frac{k+12}{4}
\end{array}$$

$u_a =$
-1 -2 -3 -4 -5 -6 -7 -8 -9 ... 2 1 0
$\mathrm{a}\mathbf{b}\mathrm{cdef}\mathbf{g}\mathrm{hijk}\mathbf{l}\mathrm{mn}\mathbf{o}\mathrm{pq}\mathbf{r}\mathrm{s}\mathbf{t}\mathrm{u}\mathbf{v}\mathrm{w}\mathbf{x}\ \dots\ \underline{\mathrm{s}}\underline{\mathbf{t}}\underline{\mathrm{u}}\underline{\mathbf{v}}\mathrm{w}\underline{\mathrm{x}}\underline{\mathrm{y}}$
-1 -2 -3 -4 -5 -6 -7 -8 ... 2 1 0
$\underline{\mathrm{z}}\mathbf{b}\mathrm{aced}\mathbf{g}\mathrm{fihj}\mathbf{l}\mathrm{km}\mathbf{o}\mathrm{np}\mathbf{r}\mathrm{q}\mathbf{t}\mathrm{s}\mathbf{v}\mathrm{u}\mathbf{x}\ \dots\ \underline{\mathrm{q}}\underline{\mathbf{t}}\underline{\mathrm{s}}\underline{\mathbf{v}}\underline{\mathrm{u}}\underline{\mathrm{w}}\underline{\mathrm{y}}$

$u_b =$
-1 -2 -3 -4 -5 -6 -7 -8 -9 ... 2 1 0
$\mathrm{a}\mathbf{b}\mathrm{cdef}\mathbf{g}\mathrm{hijk}\mathbf{l}\mathrm{mn}\mathbf{o}\mathrm{pq}\mathbf{r}\mathrm{s}\mathbf{t}\mathrm{u}\mathbf{v}\mathrm{w}\mathbf{x}\ \dots\ \underline{\mathrm{s}}\underline{\mathbf{t}}\underline{\mathrm{u}}\underline{\mathbf{v}}\mathrm{w}\underline{\mathrm{x}}\underline{\mathrm{y}}$
-1 -2 -3 -4 -5 -6 -7 -8 ... 2 1 0
$\underline{\mathrm{z}}\mathbf{b}\mathrm{aced}\mathbf{g}\mathrm{fihj}\mathbf{l}\mathrm{km}\mathbf{o}\mathrm{np}\mathbf{r}\mathrm{q}\mathbf{t}\mathrm{s}\mathbf{v}\mathrm{u}\mathbf{x}\ \dots\ \underline{\mathrm{q}}\underline{\mathbf{t}}\underline{\mathrm{s}}\underline{\mathbf{v}}\underline{\mathrm{u}}\underline{\mathrm{x}}\underline{\mathrm{w}}$

Case $k \equiv 1 \mod 8$, $\Delta = -1$:

$$\begin{array}{rcl}
\sigma_a &=& (k{-}3 \quad k{-}4 \quad ..(-4).. \quad 21 \quad 17 \quad 13 \quad 10 \quad 8 \quad 5 \quad 4 \quad 0) \\
 && (k{-}2 \quad k{-}6 \quad ..(-4).. \quad 19 \quad 15 \quad 12 \quad 9 \quad 7 \quad 3 \quad 2 \quad k{-}1) \\
\sigma_b &=& (k{-}2 \quad k{-}4 \quad ..(-4).. \quad 21 \quad 17 \quad 13 \quad 10 \quad 8 \quad 5 \quad 4 \quad 0) \\
 && (k{-}3 \quad k{-}6 \quad ..(-4).. \quad 19 \quad 15 \quad 12 \quad 9 \quad 7 \quad 3 \quad 2 \quad k{-}1) \\
I &=& \{1,6,11,14,16,..(2)..,k{-}5\} \quad K = \frac{k+11}{4}
\end{array}$$

$u_a =$
-1 -2 -3 -4 -5 -6 -7 -8 ... 2 1 0
$\mathrm{a}\mathbf{b}\mathrm{cdef}\mathbf{g}\mathrm{hijk}\mathbf{l}\mathrm{mn}\mathbf{o}\mathrm{p}\mathbf{q}\mathrm{r}\mathbf{s}\mathrm{t}\mathbf{u}\ \dots\ \underline{\mathrm{s}}\underline{\mathbf{t}}\underline{\mathrm{u}}\underline{\mathbf{v}}\mathrm{w}\underline{\mathrm{x}}\underline{\mathrm{y}}$
-1 -2 -3 -4 -5 -6 -7 ... 2 1 0
$\underline{\mathrm{z}}\mathbf{b}\mathrm{aced}\mathbf{g}\mathrm{fihj}\mathbf{l}\mathrm{km}\mathbf{o}\mathrm{n}\mathbf{q}\mathrm{p}\mathbf{s}\mathrm{r}\mathbf{u}\ \dots\ \underline{\mathrm{q}}\underline{\mathbf{t}}\underline{\mathrm{s}}\underline{\mathbf{v}}\underline{\mathrm{u}}\underline{\mathrm{w}}\underline{\mathrm{y}}$

$u_b =$
-1 -2 -3 -4 -5 -6 -7 -8 ... 2 1 0
$\mathrm{a}\mathbf{b}\mathrm{cdef}\mathbf{g}\mathrm{hijk}\mathbf{l}\mathrm{mn}\mathbf{o}\mathrm{p}\mathbf{q}\mathrm{r}\mathbf{s}\mathrm{t}\mathbf{u}\ \dots\ \underline{\mathrm{s}}\underline{\mathbf{t}}\underline{\mathrm{u}}\underline{\mathbf{v}}\mathrm{w}\underline{\mathrm{x}}\underline{\mathrm{y}}$
-1 -2 -3 -4 -5 -6 -7 ... 2 1 0
$\underline{\mathrm{z}}\mathbf{b}\mathrm{aced}\mathbf{g}\mathrm{fihj}\mathbf{l}\mathrm{km}\mathbf{o}\mathrm{n}\mathbf{q}\mathrm{p}\mathbf{s}\mathrm{r}\mathbf{u}\ \dots\ \underline{\mathrm{q}}\underline{\mathbf{t}}\underline{\mathrm{s}}\underline{\mathbf{v}}\underline{\mathrm{u}}\underline{\mathrm{x}}\underline{\mathrm{w}}$

Case $k \equiv 2 \mod 8$, $\Delta = -1$:

$$\begin{array}{rcl}
\sigma_a &=& (k{-}3 \quad k{-}4 \quad ..(-4).. \quad 18 \quad 14 \quad 10 \quad 8 \quad 5 \quad 4 \quad 0) \\
 && (k{-}2 \quad k{-}6 \quad ..(-4).. \quad 16 \quad 12 \quad 9 \quad 7 \quad 3 \quad 2 \quad k{-}1) \\
\sigma_b &=& (k{-}2 \quad k{-}4 \quad ..(-4).. \quad 18 \quad 14 \quad 10 \quad 8 \quad 5 \quad 4 \quad 0) \\
 && (k{-}3 \quad k{-}6 \quad ..(-4).. \quad 16 \quad 12 \quad 9 \quad 7 \quad 3 \quad 2 \quad k{-}1) \\
I &=& \{1,6,11,13,..(2)..,k{-}5\} \quad K = \frac{k+10}{4}
\end{array}$$

$u_a =$
-1 -2 -3 -4 -5 -6 -7 ... 2 1 0
$\mathrm{a}\mathbf{b}\mathrm{cdef}\mathbf{g}\mathrm{hijk}\mathbf{l}\mathrm{m}\mathbf{n}\mathrm{o}\mathbf{p}\mathrm{q}\mathbf{r}\ \dots\ \underline{\mathrm{s}}\underline{\mathbf{t}}\underline{\mathrm{u}}\underline{\mathbf{v}}\mathrm{w}\underline{\mathrm{x}}\underline{\mathrm{y}}$
-1 -2 -3 -4 -5 -6 ... 2 1 0
$\underline{\mathrm{z}}\mathbf{b}\mathrm{aced}\mathbf{g}\mathrm{fihj}\mathbf{l}\mathrm{k}\mathbf{n}\mathrm{m}\mathbf{p}\mathrm{o}\mathbf{r}\ \dots\ \underline{\mathrm{q}}\underline{\mathbf{t}}\underline{\mathrm{s}}\underline{\mathbf{v}}\underline{\mathrm{u}}\underline{\mathrm{w}}\underline{\mathrm{y}}$

$u_b =$
-1 -2 -3 -4 -5 -6 -7 ... 2 1 0
$\mathrm{a}\mathbf{b}\mathrm{cdef}\mathbf{g}\mathrm{hijk}\mathbf{l}\mathrm{m}\mathbf{n}\mathrm{o}\mathbf{p}\mathrm{q}\mathbf{r}\ \dots\ \underline{\mathrm{s}}\underline{\mathbf{t}}\underline{\mathrm{u}}\underline{\mathbf{v}}\mathrm{w}\underline{\mathrm{x}}\underline{\mathrm{y}}$
-1 -2 -3 -4 -5 -6 ... 2 1 0
$\underline{\mathrm{z}}\mathbf{b}\mathrm{aced}\mathbf{g}\mathrm{fihj}\mathbf{l}\mathrm{k}\mathbf{n}\mathrm{m}\mathbf{p}\mathrm{o}\mathbf{r}\ \dots\ \underline{\mathrm{q}}\underline{\mathbf{t}}\underline{\mathrm{s}}\underline{\mathbf{v}}\underline{\mathrm{u}}\underline{\mathrm{x}}\underline{\mathrm{w}}$

Case $k \equiv 3 \mod 8$, $\Delta = -1$:

$$\begin{array}{rcl}
\sigma_a & = & (k{-}3 \quad k{-}4 \quad ..(-4).. \quad 19 \quad 15 \quad 11 \quad 8 \quad 5 \quad 4 \quad 0) \\
 & & (k{-}2 \quad k{-}6 \quad ..(-4).. \quad 17 \quad 13 \quad 10 \quad 7 \quad 3 \quad 2 \quad k{-}1) \\
\sigma_b & = & (k{-}2 \quad k{-}4 \quad ..(-4).. \quad 19 \quad 15 \quad 11 \quad 8 \quad 5 \quad 4 \quad 0) \\
 & & (k{-}3 \quad k{-}6 \quad ..(-4).. \quad 17 \quad 13 \quad 10 \quad 7 \quad 3 \quad 2 \quad k{-}1) \\
I & = & \{1, 6, 9, 12, 14, ..(2).., k{-}5\} \quad K = \frac{k+9}{4}
\end{array}$$

u_a =
-1 -2 -3 -4 -5 -6 -7 2 1 0
a**b**cde f**g**h i **j** k l**mn**o**p**q r **s** . . . s **t** u**v**wxy
-1 -2 -3 -4 -5 -6 2 1 0
z**b**a c e d**g** f h **j** i k**ml** **o**n**q**p **s** . . . q **t** s **v** uwy

u_b =
-1 -2 -3 -4 -5 -6 -7 2 1 0
a**b**cde f**g**h i **j** k l**mn**o**p**q r **s** . . . s **t** u**v**wxy
-1 -2 -3 -4 -5 -6 2 1 0
z**b**a c e d**g** f h **j** i k**ml** **o**n**q**p **s** . . . q **t** s **v** uxw

Case $k \equiv 4 \mod 8$, $\Delta = -1$:

$$\begin{array}{rcl}
\sigma_a & = & (k{-}3 \quad k{-}4 \quad ..(-4).. \quad 16 \quad 12 \quad 8 \quad 5 \quad 4 \quad 0) \\
 & & (k{-}2 \quad k{-}6 \quad ..(-4).. \quad 14 \quad 10 \quad 7 \quad 3 \quad 2 \quad k{-}1) \\
\sigma_b & = & (k{-}2 \quad k{-}4 \quad ..(-4).. \quad 16 \quad 12 \quad 8 \quad 5 \quad 4 \quad 0) \\
 & & (k{-}3 \quad k{-}6 \quad ..(-4).. \quad 14 \quad 10 \quad 7 \quad 3 \quad 2 \quad k{-}1) \\
I & = & \{1, 6, 9, 11, ..(2).., k{-}5\} \quad K = \frac{k+8}{4}
\end{array}$$

u_a =
-1 -2 -3 -4 -5 -6 2 1 0
a**b**cde f**g**h i **j** k **l**m**n**o**p** . . . s **t** u**v**wxy
-1 -2 -3 -4 -5 2 1 0
z**b**a c e d**g** f h **j** i **l** k**n**m**p** . . . q **t** s **v** uwy

u_b =
-1 -2 -3 -4 -5 -6 2 1 0
a**b**cde f**g**h i **j** k **l**m**n**o**p** . . . s **t** u**v**wxy
-1 -2 -3 -4 -5 2 1 0
z**b**a c e d**g** f h **j** i **l** k**n**m**p** . . . q **t** s **v** uxw

Case $k \equiv 5 \mod 8$, $\Delta = -1$:

$$\begin{array}{rcl}
\sigma_a & = & (k{-}3 \quad k{-}4 \quad ..(-4).. \quad 9 \quad 5 \quad 4 \quad 0) \\
 & & (k{-}2 \quad k{-}6 \quad ..(-4).. \quad 7 \quad 3 \quad 2 \quad k{-}1) \\
\sigma_b & = & (k{-}2 \quad k{-}4 \quad ..(-4).. \quad 9 \quad 5 \quad 4 \quad 0) \\
 & & (k{-}3 \quad k{-}6 \quad ..(-4).. \quad 7 \quad 3 \quad 2 \quad k{-}1) \\
I & = & \{1, 6, ..(2).., k{-}5\} \quad K = \frac{k+7}{4}
\end{array}$$

u_a =
-1 -2 -3 -4 2 1 0
a**b**cde f**g**h **i** j **k** . . . s **t** u**v**wxy
-1 -2 -3 -4 2 1 0
z**b**a c e d**g** f **i** h**k** . . . q **t** s **v** uwy

u_b =
-1 -2 -3 -4 2 1 0
a**b**cde f**g**h **i** j **k** . . . s **t** u**v**wxy
-1 -2 -3 -4 2 1 0
z**b**a c e d**g** f **i** h**k** . . . q **t** s **v** uxw

Case $k \equiv 6 \mod 8$, $\Delta = -1$:

$$\begin{array}{rcllllllllllll}
\sigma_a & = & (k{-}3 & k{-}4 & ..(-4).. & 26 & 22 & 18 & 15 & 13 & 10 & 8 & 5 & 4 \;\; 0) \\
 & & (k{-}2 & k{-}6 & ..(-4).. & 24 & 20 & 17 & 14 & 12 & 9 & 7 & 3 & 2 \;\; k{-}1) \\
\sigma_b & = & (k{-}2 & k{-}4 & ..(-4).. & 26 & 22 & 18 & 15 & 13 & 10 & 8 & 5 & 4 \;\; 0) \\
 & & (k{-}3 & k{-}6 & ..(-4).. & 24 & 20 & 17 & 14 & 12 & 9 & 7 & 3 & 2 \;\; k{-}1) \\
I & = & \multicolumn{12}{l}{\{1,6,11,16,19,21,..(2)..,k{-}5\} \quad K = \frac{k+14}{4}}
\end{array}$$

$u_a =$
-1 -2 -3 -4 -5 -6 -7 -8 -9 -10 2 1 0
$a\mathbf{b}cde f\mathbf{g}h i j k \mathbf{l}mn o p\mathbf{q}r s \mathbf{t}u\mathbf{v}w\mathbf{x}y\mathbf{z} \dots \underline{s}\,\underline{\mathbf{t}}\,\underline{u}\,\underline{\mathbf{v}}\,\underline{w}\,\underline{x}\,\underline{y}$
-1 -2 -3 -4 -5 -6 -7 -8 -9 2 1 0
$\underline{z}\mathbf{b}ace d\mathbf{g}f i h j \mathbf{l}k nmo\mathbf{q}p r \mathbf{t}s\mathbf{v}u\mathbf{x}w\mathbf{z} \dots \underline{q}\,\underline{\mathbf{t}}\,\underline{s}\,\underline{\mathbf{v}}\,\underline{u}\,\underline{w}\,\underline{y}$

$u_b =$
-1 -2 -3 -4 -5 -6 -7 -8 -9 -10 2 1 0
$a\mathbf{b}cde f\mathbf{g}h i j k \mathbf{l}mn o p\mathbf{q}r s \mathbf{t}u\mathbf{v}w\mathbf{x}y\mathbf{z} \dots \underline{s}\,\underline{\mathbf{t}}\,\underline{u}\,\underline{\mathbf{v}}\,\underline{w}\,\underline{x}\,\underline{y}$
-1 -2 -3 -4 -5 -6 -7 -8 -9 2 1 0
$\underline{z}\mathbf{b}ace d\mathbf{g}f i h j \mathbf{l}k nmo\mathbf{q}p r \mathbf{t}s\mathbf{v}u\mathbf{x}w\mathbf{z} \dots \underline{q}\,\underline{\mathbf{t}}\,\underline{s}\,\underline{\mathbf{v}}\,\underline{u}\,\underline{x}\,\underline{w}$

Case $k \equiv 7 \mod 8$, $\Delta = -1$:

$$\begin{array}{rcllllll}
\sigma_a & = & (k{-}2 & 1 & 3 & 6 & ..(4).. & k{-}9 \;\; k{-}5) \\
 & & (k{-}1 & 2 & 4 & 8 & ..(4).. & k{-}7 \;\; k{-}3) \\
\sigma_b & = & (k{-}1 & 1 & 3 & 6 & ..(4).. & k{-}9 \;\; k{-}5) \\
 & & (k{-}2 & 2 & 4 & 8 & ..(4).. & k{-}7 \;\; k{-}3) \\
I & = & \multicolumn{6}{l}{\{1,6,11,16,18,..(2)..,k{-}5\} \quad K = \frac{k+13}{4}}
\end{array}$$

$u_a =$
-1 -2 -3 -4 -5 -6 -7 -8 -9 2 1 0
$a\mathbf{b}cde f\mathbf{g}h i j k \mathbf{l}mn o p\mathbf{q}r\mathbf{s}t\mathbf{u}v\mathbf{w} \dots \underline{s}\,\underline{\mathbf{t}}\,\underline{u}\,\underline{\mathbf{v}}\,\underline{w}\,\underline{x}\,\underline{y}$
-1 -2 -3 -4 -5 -6 -7 -8 2 1 0
$\underline{z}\mathbf{b}ace d\mathbf{g}f i h j \mathbf{l}k nmo\mathbf{q}p\mathbf{s}r\mathbf{u}t\mathbf{w} \dots \underline{q}\,\underline{\mathbf{t}}\,\underline{s}\,\underline{\mathbf{v}}\,\underline{u}\,\underline{w}\,\underline{y}$

$u_b =$
-1 -2 -3 -4 -5 -6 -7 -8 -9 2 1 0
$a\mathbf{b}cde f\mathbf{g}h i j k \mathbf{l}mn o p\mathbf{q}r\mathbf{s}t\mathbf{u}v\mathbf{w} \dots \underline{s}\,\underline{\mathbf{t}}\,\underline{u}\,\underline{\mathbf{v}}\,\underline{w}\,\underline{x}\,\underline{y}$
-1 -2 -3 -4 -5 -6 -7 -8 2 1 0
$\underline{z}\mathbf{b}ace d\mathbf{g}f i h j \mathbf{l}k nmo\mathbf{q}p\mathbf{s}r\mathbf{u}t\mathbf{w} \dots \underline{q}\,\underline{\mathbf{t}}\,\underline{s}\,\underline{\mathbf{v}}\,\underline{u}\,\underline{x}\,\underline{w}$

Table 3.2 – σ_a, σ_b, I, u_a and u_b. "x ..(y).. z" means every integer between x and z, with a step of y. In u_a and u_b, letters $a, b, c, \dots$ represent letters $0, 1, 2 \dots$ in Σ_k, letters $\underline{z}, \underline{y}, \underline{x}, \dots$ represent letters $k{-}1, k{-}2, k{-}3, \dots$ and letters in bold are letters in I. The over-brackets give the ODCC couples, with their number (modulo K).

- *The distance in* $\mathbf{w}$ *of two consecutive occurrences of* i *is* $\frac{s}{2} \in \{k-1, k+1\}$.
- *The distance in* $\mathbf{w}$ *of two consecutive occurrences of* j *is* $k-1$, k *or* $k+1$.
- *The distance in* $\mathbf{w}$ *between two occurrences of* ij *is at least* $\frac{(k+1)(k-1)}{2}$.
- *The distance in* $\mathbf{w}$ *between two occurrences of* ji *is at least* $\frac{(k+1)(k-1)}{2}$.

Proof. The distance between two consecutive occurrences of the same letter is $k-1$, k or $k+1$, by definition of M_k. Moreover, the distance between two consecutive occurrences of i is $\frac{s}{2}$ by the previous remark on $\mathcal{I}$. Now suppose $\Delta = 1$ (the case $\Delta = -1$ is similar). Suppose that ij appears in $\mathbf{w}$ at position x, and let $y > x$ be the next occurrence of ij in $\mathbf{w}$. The next occurrence of j cannot be at position $x+k+2$ otherwise $\mathbf{w}$ would have an forbidden repetition of excess 2. So the next occurrence of j is at position $x+k$. The gap between the q-th next occurrence of i after x and the q-th occurrence of j after $x+1$ cannot decrease, since the is are spaced by $k+1$. Suppose that there are n occurrences of i between x and y (including positions x and y) and m occurrences of j between $x+1$ and $y+1$. One has $m > n$. Then $y - x = (k+1)(n-1)$ and $y - x \geq (k-1)\cdot(m-1) \geq (k-1)n$. So $n \geq \frac{k+1}{2}$, and $y - x \geq \frac{(k+1)(k-1)}{2}$. The distance in $\mathbf{w}$ between two occurrences of ji is also at least $\frac{(k+1)(k-1)}{2}$, by symmetry. $\square$

Lemma 3.14. *l is a multiple of* $\frac{s}{2}$.

Proof. Since $|e| \geq 5$, by Fact 3.12, e contains a $i \in I$, which are spaced by $\frac{s}{2}$. $\square$

Lemma 3.15. $|e| \geq \frac{k+1}{2}$.

Proof. Since $|e| \geq 5$, by Fact 3.12, e contains a factor ij or ji for a $i \in I$ and $j \in \{0, \dots k-1\} \setminus I$. By Lemma 3.13, $|p| \geq \frac{(k-1)(k+1)}{2}$. Thus $|e| > \frac{k+1}{2}$. $\square$

Fact 3.16. For every $x, y \in \mathcal{J} \cap \{0, \frac{s}{2} - 1\}$ such that $x \neq y$ and $\gamma(x) = \gamma(y)$, we have $\gamma(x + \frac{s}{2}) \neq \gamma(y + \frac{s}{2})$.

Lemma 3.17. *l is a multiple of s.*

Proof. We know that l is a multiple of $\frac{s}{2}$. Suppose that it is not a multiple of s. Since $|e| > \frac{k+1}{2} \geq 16$, there are $x, y \in \{t, \dots t + |e| - 1\}$ such that $x, y \in \mathcal{J}$, $y - x \in \{1, 2\}$, $\gamma(x) = \gamma(y)$ and $\lfloor \frac{x}{s} \rfloor = \lfloor \frac{y}{s} \rfloor$. Thus, we have a contradiction with Fact 3.16. $\square$

Lemma 3.18. *l is a multiple of $s \cdot K$.*

Proof. Suppose that $l = n \cdot s$. Again, e contains an occurrence $x \in \mathcal{J}$ of a $j \notin I$. We have $\gamma(x) \equiv \gamma(x + s \cdot n) \mod K$. Thus by Proposition 3.11, $n \equiv 0 \mod K$. $\square$

Lemma 3.19. *If $\mathbf{u}$ is $(000, 111)$-free, then $|p| > s \cdot K$.*

Proof. Suppose that $|p| = s \cdot K$. Since $|e| \geq \frac{k}{2}$, e contains two occurrences x and y such that $x, y \in \mathcal{J}$, $\lfloor \frac{x}{s} \rfloor = \lfloor \frac{y}{s} \rfloor$, and $\gamma(y) \equiv \gamma(x) + 1 \mod K$. Let $v = \mathbf{u}\left[\lfloor \frac{x}{s} \rfloor - K : \lfloor \frac{x}{s} \rfloor - 1\right]$. Couples i and $i+1$ are good on the same time on v, which is impossible by Proposition 3.6 since u is $(000, 111)$-free. $\square$

3.3.2 Avoiding short kernel repetition

From now on, $\mathbf{w}' = h(w_{TM})$ and $\mathbf{w} = M_k(\mathbf{w}')$. If $\mathbf{w}$ has a forbidden repetition, then by results in the previous section, its period is a multiple of $s{\cdot}K$ and is at least $2{\cdot}s{\cdot}K$. In this case, its excess is at least $4{\cdot}K \geq k$, since $\frac{s}{2} \geq k-1$ and $K \geq \frac{k}{4}$. Thus this repetitions is a kernel repetition, and $\mathbf{w}'$ has a forbidden φ-kernel repetition. Let l be the least integer such that $\mathbf{w}'$ has a φ-kernel repetition (p', e') of period l and with $\frac{|e'|+k-1}{l} > \frac{1}{k-1}$. Let t be the least integer such that $\mathbf{w}'[t:\infty]$ has $p'e'$ as prefix. We suppose w.l.o.g. that the excess is maximal, that is $(p', \mathbf{w}'[t : t+|e'|])$ is not a repetition. Note that, by construction of h and by minimality of t, t is a multiple of $\frac{s}{2}$.

We now denote by w'_{TM} the infinite binary word defined as the fixed point of the following morphism :

$$\mu : \begin{cases} 0 & \mapsto 1010 \\ 1 & \mapsto 1011. \end{cases}$$

This word is the derivative word of w_{TM}, and it can also be constructed as a Toeplitz word for the pattern $101\bullet$, that is $w'_{TM} = \lim_{n\to\infty} T_k$ with $T_0 = \bullet^\omega$, and $T_{i+1} = F(T_i)$, where $F(w)$ is the word obtained from $(101\bullet)^\omega$ by replacing the sequence of all occurrences of $\bullet$ by w [36].

For $n \geq 1$ and $\delta \geq 1$, let $\chi_{(n,\delta)}$ be the infinite word such that for every $x \in \mathbb{N}$ (the sum is over GF(2)) :

$$\chi_{(n,\delta)}[x] = \sum_{i=0}^{n-1} w'_{TM}[x + i{\cdot}\delta]$$

Let $f_0(n)$ be the least r such that for every odd δ, $\chi_{(n,\delta)}$ is 0^r-free, and let $f_1(n)$ be the least r such that for every odd δ, $\chi_{(n,\delta)}$ is 1^r-free. The following comes from easy observations.

Proposition 3.20. *Let δ be odd and $x > 0$. Then*

- $f_0(1) = 2$ *and* $f_1(1) = 4$,
- $f_0(2) \leq 3$,
- $f_0(3) \leq 8$ *and* $f_1(3) \leq 4$,
- $f_0(4x) \leq 4f_0(x)$ *and* $f_1(4x) \leq 4f_1(x)$,
- $f_0(4x+1) \leq 4f_0(x)$,
- $f_0(4x+2) \leq 4f_1(x+1)$ *and* $f_0(4x+2) \leq 4f_1(x)$,
- $f_0(4x+3) \leq 4f_0(x)$.

Proof. Since w'_{TM} is $(00, 1111)$-free and contains 0 and 111, and $\chi_{(2,\delta)} \in (11BB)^\omega$, we have $f_0(1) = 2$, $f_1(1) = 4$ and $f_0(2) \leq 3$.

Suppose now that $\delta \equiv 1 \mod 4$. The case $\delta \equiv 3 \mod 4$ is proved similarly. We use the Toeplitz definition of w'_{TM}. Let $w_{(n,d)}[i] = \chi_{(n,\delta)}[4i + d]$ for every $d \in \{0, 1, 2, 3\}$, $i \geq 0$ and $n > 0$. We have $w_{(3,3)} = w_{(4,3)} = w'_{TM}$, and 0^8 cannot appear in $\chi_{(3,\delta)}$ (resp. 1^{16} cannot appear in $\chi_{(4,\delta)}$) because w'_{TM} is 0^2-free (resp 1^4-free). Moreover $w_{(3,0)} = 0^\omega$, so $\chi_{(3,\delta)}$ is 1^4-free. Similarly, other results follow from facts that:

- $w_{(4n,1)} = w_{(4n+1,1)} = \chi_{(n,\delta)}[\delta+1:\infty]$.
- $w_{(4n+2,0)} = \chi_{(n,\delta)}$.
- $w_{(4n,2)} = \overline{\chi_{(n,\delta)}[\frac{\delta+1}{2}:\infty]}$.
- $w_{(4n+2,0)} = \overline{\chi_{(n,\delta)}}$.

$\square$

By previous observations, we have:

Proposition 3.21. *For every $x \in \{1,\dots 17\}$, $f_0(x) \leq 32$.*

One can note than more generally, we have $\max(f_0(x), f_1(x)) \leq 4x$ for every $x \geq 1$. By Proposition 3.21, we have:

Proposition 3.22. *For every $n \leq 17$ and every odd $\delta \geq 33$, $\chi_{(n,\delta)}$ has no factor $0^{\delta-1}$.*

The least integer r such that $\chi_{(n,\delta)}$ is 0^r-free, for a fixed n and δ, can be found by a computer : $\mu^i(1)$ contains every factor of w'_{TM} of size at most $4^{(i-1)}+1$, so every factor of size n of w'_{TM} appear in its prefix of size $4n$. Thus the prefix of $\chi_{(n,\delta)}$ of size $4(l+(n-1)\delta)$ contains every prefix of $\chi_{(n,\delta)}$ of size at most l. Computer checks show that:

Proposition 3.23.

(i) For every $n \leq 17$ and every odd $11 \leq \delta \leq 31$, if $\chi_{(n,\delta)}$ has a factor $0^{\delta-1}$, then $(n,\delta) \in \{15,17\} \times \{11,15,17\}$.

(ii) w_{TM} has no factor w of the size $11{\cdot}n$ such that $\varphi'(w) = \mathrm{Id}_k$, for $n \in \{15,17\}$.

(iii) The maximal excess of a φ'-kernel repetition of period $n{\cdot}K$ in w_{TM}, for $n \leq 17$ and $11 \leq K$, is one.

Lemma 3.24. *If $K \geq 11$, then $l \geq 18{\cdot}s{\cdot}K$.*

Proof. Suppose that $K \geq 11$ and $l = n{\cdot}s{\cdot}K$ with $n \leq 17$.

Case 1: s divides t. Since $s{\cdot}K$ divides l, w_{TM} has a φ'-kernel repetition (p'', e'') with $p' = h(p'')$ and $|e''| \geq \left\lfloor \frac{|e'|}{s} \right\rfloor$. By Proposition 3.6 $\chi_{(n,K)}$ contains the factor 0^{K-1}. By Proposition 3.23 (i), $K \in \{11, 15, 17\}$, and by (ii), $K \in \{15, 17\}$. On the other hand, we have $|e'| \geq 3$, and we get the contradiction with (iii).

Case 2: s does not divide t. Then $t \equiv \frac{s}{2} \bmod s$. Note that in this case $n \in \{2,3\}$, otherwise the excess of (p', e') would be more than $\frac{s}{2}$, and t would not be minimal. We have $p' = \kappa_a h(w) \iota_b$ or $p' = \kappa_b h(w) \iota_a$, where w has size $K-1$. We suppose $p' = \kappa_a h(w) \iota_b$ (the other case is similar). Then w is followed by b in w_{TM}. Let $p_1 = h(w)\iota_b\kappa_a$ and $p_2 = h(w)\iota_b\kappa_b = h(wb)$. Then $\varphi(p_1) = \mathrm{Id}_k$ and $\varphi'(wb) = \varphi(p_2) = \varphi(p_1)\varphi(10)^{-1}\varphi(01) = (k{-}3 \quad k{-}1 \quad k{-}2)$. Thus for every $i \in \{0, \dots k{-}4\}$, $\varphi'(wb)[i] = i$. $\chi_{(n,K)}$ has a factor 0^{K-1}, which is impossible by Proposition 3.23 (i). $\square$

3.3.3 Avoiding long kernel repetitions

We know by the previous lemmas that $l \geq 18 \cdot s \cdot K$, and thus $|e'| \geq 8(k+1) \geq 4s$.

Since l and t are multiples of s, and by maximality of e', w_{TM} has a φ'-kernel repetition (p_1, e_1) such that $p' = h(p_1)$ and $e' = h(e_1)L$. Thus:

$$\frac{|e_1| \cdot s + \ell + k - 1}{|p_1| \cdot s} = \frac{|e'| + k - 1}{l} > \frac{1}{k-1}$$

Note that $|e_1| \geq 4$. The definition of markable words are those of Chapter 2. By Corollary 2.10, and since every factor of w_{TM} of size at least 4 is (ν_{TM}, w_{TM})-markable, $h(w_{TM})$ has a φ'-kernel repetition (p_2, e_2) with $p_1 = \nu_{TM}(p_2)$ and $e_1 = \nu_{TM}(e_2)$. Thus $\mathbf{w}$ has a repetition of exponent E, with :

$$E = \frac{|e_2| \cdot s + \ell + k - 1}{|p_2| \cdot s} > \frac{|e_1| \cdot s + \ell + k - 1}{|p_1| \cdot s} > \frac{1}{k-1}$$

which is a forbidden repetition of period $\frac{l}{2}$. We have a contradiction.

3.4 Alphabets with less than 9 letters

Ochem's conjecture is already proved for several cases. Chalopin and Ochem proved the first case of the conjecture for 5 letters, and the second case for 6 letters [51]. The cases $9 \leq k \leq 38$ are proved in Chapter 2. Theorem 3.7 prove it for $k \geq 24$. We give here constructions for the last cases.

All the words follow the same construction. Their Pansiot code is the morphic word $h(g^\infty(a))$, where h and g are given in the following table ($\Delta = 1$ for first case of the conjecture, and $\Delta = -1$ for the second case).

k	Δ	g	h	x_h	x_g
5	1	$a \to abaababaababababaabaab$ $aababababaababaabaabab$ $b \to abaabaabaababaababaa$ $baabaabababaababaaba$	$a \to 101101$ $b \to 101010$	13	40
6	-1	$a \to aaab$ $b \to bbba$	$a \to 1010110110$ $b \to 1011010101$	39	5
6	1	$a \to abaabb$ $b \to abbaba$	$a \to 101010110101010110101$ $b \to 101010110101101010110$	46	12
7	-1	$a \to ababa$ $b \to bbabb$	$a \to 1010101011011010101010 10$ $b \to 101010101101101101101101$	24	7
7	1	$a \to aaab$ $b \to bbba$	$a \to 1011010110110110$ $b \to 1011011010110101$	63	5
8	-1	$a \to abbba$ $b \to baaab$	$a \to 101010101101010110101$ $b \to 101010110101101010110$	30	5
8	1	$a \to aab$ $b \to baa$	$a \to 1011011011010101101010101 01$ $b \to 101101101101101010101011010$	27	7

Table 3.3 – Construction for cases $k \leq 8$.

The proof also follows ideas of Chapter 2. Let $\mathbf{w}' = g^\infty(a)$ and $\mathbf{w} = h(\mathbf{w}')$ for one case in Table 3.3. By Remark 3.2, $M_k(w)$ has at least one letter of frequency $\frac{1}{k+\Delta}$. We show that $M_k(w)$ is a Dejean word. Let $s = |h(a)| = |h(b)|$ and $s' = |g(a)| = |g(b)|$. Let ℓ (resp. ℓ') be the size of the largest common prefix of $h(a)$ and $h(b)$ (resp. $g(a)$ and $g(b)$). Let $\sigma_a = \varphi(h(a))$ and $\sigma_b = \varphi(h(b))$. Let $\varphi' : \{a,b\}^* \to \mathbb{S}_k$ be the morphism such that $\varphi'(a) = \sigma_a$ and $\varphi'(b) = \sigma(b)$. One can easily check by computer that:

Fact 3.25. There is a $\sigma \in \mathbb{S}_k$ such that for every $x \in \{a,b\}$, $\varphi'(g(x)) = \sigma \cdot \varphi'(x) \cdot \sigma^{-1}$.

Let x_h and x_h defined in Table 3.3. One can check by computer that:

Fact 3.26.

- Every factor w of $\mathbf{w}$ such that $|w| \geq x_h$ is $(h, \mathbf{w}')$-markable.
- Every factor w of $\mathbf{w}'$ such that $|w'| \geq x_g$ is $(g, \mathbf{w}')$-markable.

A Ψ-kernel repetition (p,e) is *weak* if $\frac{|e|+\frac{6}{5}}{|p|} \geq \frac{1}{k-1}$. The following corollary follow from Lemma 2.5, Fact 3.25 and Fact 3.26.

Corollary 3.27.

- *If* $\mathbf{w}$ *has a* φ*-kernel-repetition* (p,e) *with* $|e| \geq x_h$*, then* $\mathbf{w}'$ *has a* φ'*-kernel-repetition* (p',e') *with* $s \cdot |e'| + \ell \geq |e|$ *and* $s \cdot |p'| = |p|$*.*
- *If* $\mathbf{w}'$ *has a* φ'*-kernel-repetition* (p,e) *with* $|e| \geq x_g$*, then* $\mathbf{w}'$ *has a* φ'*-kernel-repetition* (p',e') *with* $s' \cdot |e'| + \ell' \geq |e|$ *and* $s' \cdot |p'| = |p|$*.*

Therefore if $M_k(\mathbf{w})$ has a forbidden kernel repetition of excess at least $k - 1 + x_h$, then $\mathbf{w}'$ has a φ'-kernel repetition (p',e') such that

$$\frac{s \cdot |e'| + k - 1 + \ell}{s \cdot |p'|} > \frac{1}{k-1}.$$

Note that (p',e') is a weak repetition since $\frac{k-1+\ell}{s} \leq \frac{6}{5}$. Similarly, if $\mathbf{w}'$ has a weak φ'-kernel repetition (p',e') of excess at least x_g, then $\mathbf{w}'$ has a φ'-kernel repetition (p'',e'') such that

$$\frac{s' \cdot |e''| + \ell' + \frac{6}{5}}{s' \cdot |p''|} > \frac{1}{k-1}.$$

Then (p'',e'') is also a weak repetition since $\frac{\ell'+\frac{6}{5}}{s'} \leq \frac{6}{5}$. The fact that $M_k(\mathbf{w})$ is a Dejean word follow from the computer checked facts that:

- $M_k(\mathbf{w})$ has no forbidden repetition of excess at most $k - 1 + x_h$.
- $\mathbf{w}'$ has no weak repetition of excess at most x_g.

Chapter 4

Finite repetition threshold

We investigate the finite repetition threshold for k-letter alphabets, $k \geq 4$, that is the smallest number r for which there exists an infinite r^+-free word containing a finite number of r-powers. We show that there exists an infinite Dejean word on a 4-letter alphabet (*i.e.* a word without factors of exponent more than $\frac{7}{5}$) containing only two $\frac{7}{5}$-powers. For a 5-letter alphabet, we show that there exists an infinite Dejean word containing only 60 $\frac{5}{4}$-powers, and we conjecture that this number can be lowered to 45. Finally, we show that the finite repetition threshold for k letters is equal to the repetition threshold for k letters, for every $k \geq 6$.

This chapter is based on paper [2] (joint work with Golnaz Badkobeh and Maxime Crochemore).

4.1 Introduction

Following the study of infinite words avoiding repetitions in relation to Dejean's statement on the repetition threshold of alphabets [64] we show that it is possible to impose more constraints on words. We are interested in infinite words whose maximal exponent of its finite factors does not exceed Dejean's threshold and that contain a finite number of factors having the maximal exponent. This introduces the notion of finite repetition threshold (see [38, 39]). Imposing this constraint is not possible on the binary alphabet whose finite repetition threshold is $\frac{7}{3}$ while the repetition threshold is 2 (see [138, 132]), but can be satisfied for the ternary alphabet [39]. We show here that the result also holds for larger alphabets. This confirms the intuition given by the growth rates of words having the smallest exponent according to their alphabet size (see [99, 140]).

Associated with the finite repetition threshold is the smallest number of factors of highest exponent that an infinite word can accommodate (see [77, 37]). We show here that there exists an infinite word on a 4-letter alphabet containing only two $\frac{7}{5}$-powers and no factor of exponent more than $\frac{7}{5}$. The only known proofs of the $\frac{7}{5}$ repetition threshold for 4 letters are due to Pansiot [125] and Rao [28]; both of their words contain 24 $\frac{7}{5}$-powers. On 5 letters, the proof of the $\frac{5}{4}$ threshold by Moulin-Ollagnier [118] provides a word with 360 $\frac{5}{4}$-powers of periods 4, 12 and 44. We show that this number can be reduced to 60 and conjecture that it can be lowered to 45, the smallest possible number.

Both results also provide in fact new proofs of the repetition thresholds for the corresponding alphabet sizes 4 and 5. The question on the smallest number of factors of highest exponent in a Dejean word remains open for larger alphabets.

4.2 Preliminaries

The *finite repetition threshold* for k letters is the smallest number $\mathrm{FRT}(k)$ for which there exists an infinite $\mathrm{FRT}(k)^+$-free word containing a finite number of $\mathrm{RT}(k)$-powers (that is, it has a finite number of limit repetitions).

It is known that any infinite $\frac{7}{3}$-free infinite binary word contains an arbitrary number of squares [138, 132]. However, there exists an infinite binary word whose maximal exponent does not exceed $\frac{7}{3}$ and all of its squares have period length at most 7. In [38], the associated minimal number of squares that an infinite binary word can accommodate is given as follows: there exists an infinite binary word containing only 12 squares whose maximal exponent is $\frac{7}{3}$. The proof is based on a HD0L-system exploiting two special non-uniform morphisms, the first one on 6-letter alphabet and the second from 6 letters to binary. Furthermore, a simple construction of all binary words with only 11 squares whose maximal exponent is $\frac{7}{3}$ showed that this set is finite and that its longest element has length 116, which shows the minimality of 12.

This idea was extended and further studied in [39] on ternary words. The result is as follows: there exists an infinite ternary Dejean word containing only two $\frac{7}{4}$-powers. The proof is based on a 160-uniform morphism which translates any infinite Dejean word on 4 letters to an infinite Dejean word on 3 letters containing only two $\frac{7}{4}$-powers.

Throughout this chapter, in order to prove the existence of an infinite word complying with some properties, the following method is used. The main technique is to design two or more morphisms generating an appropriate infinite binary word and then translate that by the inverse of the Pansiot coding. One of the experimental techniques that we used consists of the following steps. We generate a long enough word satisfying the pre-defined constraints using a backtracking strategy, and we translate this word to a binary word by applying the Pansiot coding. Then, we search for its most repetitive motifs, and using selective elements of the set of motifs, we try to decode the word to find its pre-image according to the morphism defined by the motifs. If necessary, we iterate the previous step with the new word (pre-image of the first word). Backtracking is a general algorithm for finding all (or some) solutions to some computational problem; it incrementally builds candidates to the solutions, and abandons each partial candidate as soon as it determines it cannot possibly be completed to a valid solution.

4.3 Finite repetition threshold for 4-letter alphabets

Since the repetition threshold for a 4-letter alphabet is $\frac{7}{5}$, it suffices to show that there exists a $\frac{7}{5}^+$-free infinite word on Σ_4 with finitely many limit repetitions (that is $\frac{7}{5}$-powers). There are two proofs of Dejean's conjecture for 4-letter alphabets, by Pansiot [125] and Rao [28]. In both cases the number of limit repetitions contained in the infinite words is 24. This proves that the finite repetition threshold for 4 letters is $\frac{7}{5}$. In this section, we prove the following:

Theorem 4.1. *The finite repetition threshold for* 4*-letter alphabets is* $\frac{7}{5}$ *and the minimal number of* $\frac{7}{5}$*-powers is* 2*.*

A computer check shows that a word on a 4-letter alphabet for which the maximal exponent of factors is $\frac{7}{5}$ and that contains at most one limit repetition has maximal length 230. Then, to prove Theorem 4.1, we give a morphic word which is the Pansiot code of a Dejean word on 4 letters with only two limit repetitions. The correctness proof follows the plan and notations introduced in [28]. However, since the morphism φ' will be simpler here, we can make the proof self-contained. Informally, the idea is to prove that if the morphic word has a forbidden repetition with a long enough period, then it has a smaller forbidden repetition. Thus it remains to prove that the morphic word has no forbidden repetition with a period bounded by a constant, which can be done by a finite case analysis. Let:

$$f : \begin{cases} \mathtt{a} & \to \mathtt{abc} \\ \mathtt{b} & \to \mathtt{cda} \\ \mathtt{c} & \to \mathtt{adc} \\ \mathtt{d} & \to \mathtt{cba} \end{cases}$$

$$g : \begin{cases} \mathtt{a} & \to \mathtt{aacbbaaccbaabcabc} \\ \mathtt{b} & \to \mathtt{aacbacbaabbcaabbc} \\ \mathtt{c} & \to \mathtt{cbaaccbbaccabcabc} \\ \mathtt{d} & \to \mathtt{aacbaccaabbcaabbc} \end{cases}$$

$$h : \begin{cases} \mathtt{a} \to & \mathtt{101101010110110101101101010110101011011010101101101010110101} \\ & \mathtt{011011010101101101010110101011011010101} \\ \mathtt{b} \to & \mathtt{101101010110110101101101010110110101011010101101101010110110} \\ & \mathtt{101011010101101101010110110101011010101} \\ \mathtt{c} \to & \mathtt{101101010110110101101101010110110101011011010101101010110110} \\ & \mathtt{101011011010101101010110110101011011010}. \end{cases}$$

The rest of this section is devoted to the proof of the following theorem.

Theorem 4.2. *$w_0 = \mathtt{M}_4(h(g(f^\infty(\mathtt{a}))))$ is $\frac{7}{5}^+$-free and it contains only two $\frac{7}{5}$-powers:* $(3421432412, 3421)$ *and* $(1423412432, 1423)$.

Remark 4.3. A computer check shows that the Pansiot code of every long enough $\frac{7}{5}^+$-free word on 4-letter alphabet with at most two limit repetitions contains $h(x)$ as factor, for an $x \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$. Moreover, every Pansiot code of a Dejean word with at most two limit repetitions starting with $h(x)$ (for $x \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$) must be followed by $h(y)$, for a $y \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$. Thus the morphism h in our construction is unavoidable, *i.e.* for every Dejean word w which proves Theorem 4.1, $P_4(w)$ must be the image by h of a ternary word (modulo the shift operation).

The following properties derive from simple observations:

- f is 3-uniform, g is 17-uniform and h is 99-uniform. Thus $h \circ g$ is 1683-uniform. (A morphism $f : \Sigma^* \to \Sigma'^*$ is *l-uniform*, $l \in \mathbb{N}$, if for every $x \in \Sigma$, $|f(x)| = l$.)

- f, g, h and $h \circ g$ are comma-free. (A morphism $f : \Sigma^* \to \Sigma'^*$ is *comma-free* if whenever $f(xy) = uf(z)v$, then either $u = \epsilon$ or $v = \epsilon$, for every $x, y, z \in \Sigma$ and $u, v \in \Sigma'^*$.)

- The longest common prefix in $\{h \circ g(\mathtt{a}), h \circ g(\mathtt{b}), h \circ g(\mathtt{c}), h \circ g(\mathtt{d})\}$ has size 635 and the longest common suffix has size 990.

- For every $x \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$, $\varphi(h(x)) = (13)$.

The last fact can be verified by a computer check (or by a tedious hand check). The notion of Ψ-kernel repetition is central in [118, 28]. However, the proof can be simplified here since $\varphi(h(x)) = (13)$ for every $x \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$ (which is not true for cases in [28]). Since g and h are uniform and of odd-size, for every $x \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}\}$, $\varphi(h(g(x))) = (13)$ and $\varphi(h(g(f(x)))) = (13)$. Let $\varphi' : \{0,1,2,3\}^* \to \mathbb{S}_4$ such that $\varphi'(u) = (13)^{|u|}$. Thus (p,q) is a φ'-kernel repetition if (p,q) is a repetition, and $|p|$ is even. The following lemma gives a relation between φ-kernel repetitions in $w_1 = h(g(f^\infty(\mathtt{a})))$ and φ'-kernel repetitions in $w_2 = f^\infty(\mathtt{a})$.

Lemma 4.4. *Let (p_1, e_1) be a φ-kernel-repetition of w_1. If $|e_1| \geq 3365$, then w_2 has a φ'-kernel-repetition (p_2, e_2) with $|e_2| \geq \left\lceil \frac{|e_1|-1625}{1683} \right\rceil$ and $|p_1| = 1683 \cdot |p_2|$.*

Proof. Suppose w.l.o.g. that (p_1, e_1) is a maximal repetition, *i.e.* there is no repetition (p_1', e_1') in w_1 such that $|p_1'| = |p_1|$ and p_1e_1 is a proper factor of $p_1'e_1'$. If $|e_1| \geq 3365 = 2 \cdot 1683 - 1$, then $h \circ g(\mathtt{a})$, $h \circ g(\mathtt{b})$ or $h \circ g(\mathtt{c})$ appears as a factor in e_1. Since $h \circ g$ is comma-free and 1683-uniform, $|p_1|$ is a multiple of 1683. Let $n \in \mathbb{N}$ such that $|p_1| = n \cdot 1683$. Then there is a factor $u = a_1 \ldots a_l$ in w_2 such that $h \circ g(u) = vp_1e_1v'$, v is a proper prefix of $h \circ g(a_1)$ and v' is a proper suffix of $h \circ g(a_l)$. Since (p_1, e_1) is a repetition of period $n \cdot 1683$, for every $n + 1 < i < l$, $a_i = a_{i-n}$. Thus $(p_2, e_2) = (a_2 \ldots a_{n+1}, a_{n+2} \ldots a_{l-1})$ is a repetition in w_2 of period n. Moreover, $\varphi'(p_2) = \varphi(h \circ g(p_2)) = \mathrm{Id}_4$ since $\varphi(p_1) = \mathrm{Id}_k$, and p_1 is conjugate to $h \circ g(p_2)$. (We recall that two words w and w' are *conjugated* if there are u and v such that $w = uv$ and $w' = vu$.) Since p_1e_1 is maximal on the left, $|v| \geq 693$, and since p_1e_1 is maximal on the right, $|v'| \geq 1048$. Thus $|e_1| - 1625 \leq 1683 \cdot |e_2|$, and w_2 has a φ'-kernel repetition (p_2, e_2) with $|e_2| \geq \left\lceil \frac{|e_1|-1625}{1683} \right\rceil$ and $|p_1| = 1683 \cdot |p_2|$. □

The proof of the following Lemma is similar, and is omitted.

Lemma 4.5. *If (p_2, e_2) is a φ'-kernel repetition of $w_2 = f^\infty(\mathtt{a})$ with $|e_2| \geq 5$, then w_2 has a φ'-kernel-repetition (p_2', e_2') with $|e_2'| \geq \left\lceil \frac{|e_2|-2}{3} \right\rceil$ and $|p_2| = 3 \cdot |p_2'|$.*

Lemma 4.6. *Suppose that w_2 has a φ'-kernel-repetition (p_2, e_2) with $|e_2| \geq 5$ and $\frac{|e_2|+1}{|p_2|} \geq \frac{2}{5}$. Then there exists a φ'-kernel-repetition (p_2', e_2') with $|p_2| = 3 \cdot |p_2'|$ and $\frac{|e_2'|+1}{|p_2'|} \geq \frac{2}{5}$.*

Proof. By Lemma 4.5,

$$\frac{2}{5} \leq \frac{|e_2|+1}{|p_2|} \leq \frac{3 \cdot |e_2'| + 3}{3 \cdot |p_2'|} = \frac{|e_2'|+1}{|p_2'|}.$$

□

The following fact can be verified by a computer check:

Fact 4.7. There is no φ'-kernel-repetition (p_2, e_2) with $2 \leq |e_2| < 5$ and $\frac{|e_2|+1}{|p_2|} \geq \frac{2}{5}$ in w_2.

Thus by Lemma 4.6:

Corollary 4.8. *There is no φ'-kernel-repetition (p_2, e_2) with $2 \leq |e_2|$ and $\frac{|e_2|+1}{|p_2|} \geq \frac{2}{5}$ in w_2.*

Lemma 4.9. *w_1 has no φ-kernel-repetition (p_1, e_1) with $|e_1| \geq 3 \cdot 1683$ and $\frac{|e_1|+3}{|p_1|} \geq \frac{2}{5}$.*

Proof. Suppose that w_1 has a φ-kernel-repetition (p_1, e_1) with $|e_1| \geq 3 \cdot 1683$ and $\frac{|e_1|+3}{|p_1|} \geq \frac{2}{5}$. By Lemma 4.4, w_2 has a φ'-kernel repetition (p_2, e_2) with $|e_2| \geq 2$ and

$$\frac{2}{5} \leq \frac{|e_1|+3}{|p_1|} \leq \frac{1683 \cdot |e_2| + 1625 + 3}{1683 \cdot |p_2|} < \frac{|e_2|+1}{|p_2|}.$$

By Corollary 4.8, w_2 has no such φ'-kernel repetition. Contradiction. $\square$

By Lemma 4.9, if w_1 has a φ-kernel repetition (p_1, e_1) with $\frac{|p_1 e_1|+3}{|p_1|} \geq \frac{7}{5}$, then $|p_1| \leq \frac{5}{2}(|e_1|+3) < \frac{5 \cdot (3 \cdot 1683 + 3)}{2}$, that is $|p_1| < 12630$. By Proposition 2.1 Lemma lm:mo, and since w_1 is the Pansiot code of w_0, w_0 has no repetition (p, e) with $|p| \geq 12633$ and $\frac{|pe|}{|p|} \geq \frac{7}{5}$. To complete the proof of Theorem 4.2, it suffices to show that for every repetition (p, e) in w_0 with $|p| < 12633$, either $\frac{|pe|}{|p|} < \frac{7}{5}$, or $\frac{|pe|}{|p|} = \frac{7}{5}$ and $(p, e) \in \{(3421432412, 3421), (1423412432, 1423)\}$. This fact has been verified by a computer check.

4.4 Finite repetition threshold for 5-letter alphabets

This section is devoted to the study of the minimal number of limit repetitions over all Dejean words on a 5-letter alphabet. Moulin-Ollagnier gave a proof of Dejean's conjecture for $k = 5$ (see [118]). Let:

$$m : \begin{cases} \mathtt{0} & \to \mathtt{0101011011010101101 10} \\ \mathtt{1} & \to \mathtt{10101010110110110110 1.} \end{cases}$$

Then $\mathtt{M}_5(m^\infty(\mathtt{0}))$ is $\frac{5}{4}^+$-free. We claim without proof that it contains 360 limit repetitions, of which a third have period 4, a third period 12 and the remaining have period 44. This proves that the finite repetition threshold for 5-letter alphabets is $\frac{5}{4}$. We show, with an explicit construction, that the number of limit repetitions can be lowered to 60, and we conjecture that the minimal number is 45. Most of the intermediate proofs are similar to those in Section 4.3, and are omitted. Let:

$$f : \begin{cases} \mathtt{a} & \to \mathtt{aaabbababbaaabbaabb} \\ \mathtt{b} & \to \mathtt{aabbbaababaabbbaabb} \end{cases}$$

$$g : \begin{cases} \mathtt{a} & \to \mathtt{aaaababbbbababaaaababbb} \\ \mathtt{b} & \to \mathtt{bbbbabaaaabababbbbabaaa} \end{cases}$$

$$h : \begin{cases} \mathtt{a} \to & \mathtt{1101101010101101101010101101101010110110101011011010 10110} \\ & \mathtt{1101101101010101101101010110110101011011011010101010110} \\ \mathtt{b} \to & \mathtt{1101101010110110101011011010101011011011011010101101101 01} \\ & \mathtt{01101101010110110101010110110101010110110110101010110.} \end{cases}$$

Let $w_2 = f^{\infty}(\mathtt{a})$, $w_1 = h(g(w_2))$ and $w_0 = \mathtt{M}_5(w_1)$.

Theorem 4.10. *w_0 is a Dejean word on 5 letters, and it contains only 60 limit repetitions, all of which have period 4.*

The following properties will help with the proof of Theorem 4.10:

- f is 19-uniform, g is 29-uniform and h is 113-uniform. Thus $h \circ g$ is 3277-uniform.
- f, g, h and $h \circ g$ are comma-free.
- The longest common prefix in $\{h \circ g(\mathtt{a}), h \circ g(\mathtt{b})\}$ has size 11 and the longest common suffix has size 24.
- For every $x \in \{\mathtt{a}, \mathtt{b}\}$, $\varphi(h(x)) = (12)(354)$, thus for every $x \in \{\mathtt{a}, \mathtt{b}\}$, $\varphi(h(g(x)) = (12)(345)$ and $\varphi(h(g(f(x))) = (12)(345)$.

Let $\varphi' : \{0, 1, 2, 3, 4\}^* \to \mathbb{S}_5$ such that $\varphi'(u) = [(12)(345)]^{|u|}$. Thus (p, q) is a φ'-kernel repetition if and only if (p, q) is a repetition, and $|p|$ is divisible by 6.

Lemma 4.11. *Let (p_1, e_1) be a φ-kernel-repetition of $w_1 = h(g(f^{\infty}(\mathtt{a})))$. If $|e_1| \geq 6553$, then $w_2 = f^{\infty}(\mathtt{a})$ has a φ'-kernel-repetition (p_2, e_2) with $|e_2| \geq \left\lceil \frac{|e_1|-35}{3277} \right\rceil$ and $|p_1| = 3277 \cdot |p_2|$.*

Lemma 4.12. *If $|e_2| \geq 37$, then $w_2 = f^{\infty}(\mathtt{a})$ has a φ'-kernel-repetition (p_2', e_2') with $|e_2'| \geq \left\lceil \frac{|e_2|-8}{19} \right\rceil$ and $|p_2| = 19 \cdot |p_2'|$.*

Here, we adapt the same approach as in Section 4.3 (Lemma 4.6 and Fact 4.7) with the appropriate changes based on the size of the morphism f and the exponent $\frac{5}{4}$. The next corollary follows:

Corollary 4.13. *There is no φ'-kernel-repetition (p_2, e_2) with $6 \leq |e_2|$ and $\frac{|e_2|+1}{|p_2|} \geq \frac{1}{4}$ in w_2.*

Lemma 4.14. *w_1 has no φ-kernel-repetition (p_1, e_1) with $|e_1| \geq 6 \cdot 3277$ and $\frac{|e_1|+4}{|p_1|} \geq \frac{1}{4}$.*

The proof of Lemma 4.14 is similar to the proof of Lemma 4.9, and is a direct consequence of Lemma 4.11 and 4.12. By Lemma 4.14, if w_1 has a φ-kernel repetition (p_1, e_1) with $\frac{|p_1 e_1|+4}{|p_1|} \geq \frac{5}{4}$, then $|p_1| \leq \frac{4}{1}(|e_1| + 4) < 4 \cdot (6 \cdot 3277 + 4)$, that is $|p_1| < 78664$. By Proposition 2.1 Lemma lm:mo, and since w_1 is the Pansiot code of w_0, w_0 has no repetition (p, e) with $|p| \geq 78664$ and $\frac{|pe|}{|p|} \geq \frac{5}{4}$. A computer check showed that among every repetition (p, e) in w_0 of period at most 78664, none has an exponent greater than $\frac{5}{4}$. This proves that $\mathrm{FRT}(5) = \frac{5}{4}$. This check also reveals that there are only 60 limit repetitions (p, e) in w_0, and for every limit repetition, $|e| = 1$. This concludes the proof of Theorem 4.10.

To conclude this section, we give lower bounds on the number of limit repetitions for a Dejean word on 5 letters. The following facts have been verified by a computer check. A standard (and easily parallelizable) backtrack algorithm written in C++ took approximately 3 days (resp. 120 days) of single-core time on a 2.1GHz CPU to verify fact (a) (resp. fact (b)).

Fact 4.15.

(a) A $\frac{5}{4}^+$-free word on a 5-letter alphabet that contains at most 44 limit repetitions has size at most 4648.

(b) A $\frac{5}{4}^+$-free word on a 5-letter alphabet that contains at most 45 limit repetitions, and such that every limit repetition has period 4, has size at most 7331.

Thus the minimal number of limit repetitions over all Dejean words on 5 letters is between 45 and 60. Based on computer experiments, we conjecture the following.

Conjecture 4.16.

- There exists an infinite Dejean word on a 5-letter alphabet with only 45 limit repetitions.
- There exists an infinite Dejean word on a 5-letter alphabet with only 46 limit repetitions, and such that every limit repetition has period 4.

4.5 Finite repetition threshold for k-letter alphabets, $k \geq 6$

Looking at the existing proofs for Dejean's conjecture shows in fact $\mathrm{FRT}(k) = \mathrm{RT}(k)$ for $k \geq 6$, that is, known constructions of Dejean words have finitely many limit repetitions.

Lemma 4.17. *For every* $5 \leq k \leq 11$, $\mathrm{FRT}(k) = \mathrm{RT}(k)$.

Proof. Moulin-Ollagnier gave uniform morphisms h_k, for $5 \leq k \leq 11$, such that $\mathtt{M}_k(h_k^\infty(1))$ is a Dejean word on a k-letter alphabet [118]. We show that these Dejean words have finitely many limit repetitions. We fix a $5 \leq k \leq 11$, and let $h = h_k$. Let $u = |h(0)| = |h(1)|$, and let L be the longest common prefix of $h(0)$ and $h(1)$. Note that the last letters of $h(0)$ and $h(1)$ differ. Suppose that $\mathtt{M}_k(h^\infty(1))$ has infinitely many limit repetitions. Let $\mathcal{L}$ be the set of φ-kernel repetitions (p, e) in $h^\infty(1)$ with $\frac{|e|+k-1}{|p|} = \frac{1}{k-1}$, that is φ-kernel repetitions which correspond to a limit repetition. Since $\mathtt{M}_k(h^\infty(1))$ has infinitely many limit repetitions, $\mathcal{L}$ is also infinite. By [118, Corollary 3.20], there is a repetition $(p, e) \in \mathcal{L}$ and a $n > 0$ such that $(h^n(p), \mu^n(e)) \in \mathcal{L}$, where $\mu(w) = h(w)L$. Then:

$$\frac{|e| + k - 1}{|p|} = \frac{u^n \cdot |e| + |L| \cdot \sum_{i=0}^{n-1} u^i + k - 1}{u^n \cdot |p|}$$

which is satisfied when:

$$(u - 1) \cdot (k - 1) = |L|.$$

We have a contradiction, since $|L| \leq u - 1$ and $k \geq 5$. □

For $k \geq 12$ we use the following lemma.

Lemma 4.18. *Let* $k \geq 5$. *Let* w_{TM} *be the Prouhet-Thue-Morse word, that is* $w_{TM} = \nu_{TM}^\infty(0)$ *where* $\nu_{TM} : 0 \to 01, 1 \to 10$. *Let* $w_1 = h(w_{TM})$ *be a binary word such that:*

1. *$w_0 = \mathtt{M}_k(w_1)$ is a Dejean word on a k-letter alphabet.*

2. *$h : \{0,1\}^* \to \{0,1\}^*$ is n-uniform,*

3. *there exists $\sigma \in \mathbb{S}_k$ such that $\varphi'(0)\varphi'(1) = \sigma\varphi'(0)\sigma^{-1}$ and $\varphi'(1)\varphi'(0) = \sigma\varphi'(1)\sigma^{-1}$, where $\varphi' : \{0,1\}^* \to \mathbb{S}_k$ is the morphism such that $\varphi'(0) = \varphi(h(0))$ and $\varphi'(1) = \varphi(h(1))$.*

Then w_0 has finitely many limit repetitions.

Proof. Note that w_1 cannot contain arbitrarily large powers, otherwise w_0 would also contain arbitrarily large powers. We have $h(0) \neq h(1)$, since the Pansiot code of a Dejean word is not periodic. Thus we can suppose w.l.o.g. that the last letters of $h(0)$ and $h(1)$ differ, otherwise we replace $h(0)$ (resp. $h(1)$) by $uh(0)u^{-1}$ (resp. $uh(1)u^{-1}$), where u is the largest common suffix of $h(0)$ and $h(1)$. Moreover we can suppose w.l.o.g. that the last letter of $h(x)$ is x for $x \in \{0,1\}$, otherwise we exchange $h(0)$ and $h(1)$ (note the factor set of w_{TM} is closed under the complementation). Let L be the largest common prefix of $h(0)$ and $h(1)$, and let ℓ be the size of L.

Claim 4.19. There is a $B \in \mathbb{N}$ such that for every φ-kernel repetition (p, e) in w_1 with $|e| \geq B$, $|p|$ is a multiple of n.

Proof. Let $\overline{v}$, where v is a binary word, be the image of v by the morphism $0 \to 1, 1 \to 0$. Let $M = \{1 \leq i \leq n : h(0)[i] = h(1)[i]\}$, $N = \{1 \leq i \leq n : h(0)[i] = 0$ and $h(1)[i] = 1\}$ and $N' = \{1 \leq i \leq n : h(0)[i] = 1$ and $h(1)[i] = 0\}$. Note that $\{M, N, N'\}$ is a partition of $\{1, \ldots, n\}$, and since the last letter of $h(x)$ is x, we have $n \in N$. Suppose that the claim is false. Then there are arbitrarily large factors u and u' of w_{TM}, with $|u| = |u'| + 1 \geq 4$, such that $vh(u')$ is a prefix of $h(u)$, where v is a non-empty proper suffix of $h(0)$ or $h(1)$. Since w_{TM} is cube-free, u', $u[1 : |u| - 1]$ and $u[2 : |u|]$ contain 0 and 1 as factors. Thus for every $i \in M$, $i + |v| \in M \pmod{n}$, that is $M + |v| = M \pmod{n}$. Since $n \in N$, we have $|v| \notin M$. Since the last letter of $h(x)$ is x (for $x \in \{0,1\}$), u' is either a suffix of u or of $\overline{u}$, depending on whether $|v| \in N$ or $|v| \in N'$. If u' is a suffix of u, then $h(u')[i] = h(u')[i + n - |v|]$ for every $1 \leq i \leq n \cdot |u'| - n + |v|$, that is $h(u')$ is a repetition of period $n - |v|$. Suppose now that u' is a suffix of $\overline{u}$. Let $1 \leq i \leq n \cdot |u'| - 2(n - |v|)$. If $i \in M \pmod{n}$, $h(u')[i] = h(u')[i + n - |v|] = h(u')[i + 2(n - |v|)]$, since $\{i, i + n - |v|]\} \subseteq M \pmod{n}$. Otherwise $h(u')[i] = \overline{h(u')[i + n - |v|]} = h(u')[i + 2(n - |v|)]$, since $\{i, i + n - |v|]\} \subseteq \{1 \ldots n\} \setminus M \pmod{n}$. Thus $h(u')$ is a repetition of period $2(n - |v|)$. In all cases, $h(u')$ is a repetition of period at most $2n$. Hence w_1 contains arbitrarily large powers, and we have a contradiction. □

Suppose that w_0 has infinitely many limit repetitions. Then w_1 has infinitely many φ-kernel repetitions (p, e) with $\frac{|e|+k-1}{|p|} = \frac{1}{k-1}$. By Claim 4.19, if e is long enough then $|p|$ is a multiple of n, and w_{TM} has a repetition (p', e') such that $n \cdot |p'| = |p|$, p is conjugated to $h(p')$ and $|e| = n \cdot |e'| + \ell$. Since (p, e) is a φ-kernel repetition, $\varphi(p) = \mathrm{Id}_k$ and $\varphi(h(p')) = \mathrm{Id}_k$. By condition (3), $\varphi'(p') = \mathrm{Id}_k$, and (p', e') is a φ'-kernel repetition of w_{TM}.

Thus the word w_{TM} has infinitely many φ'-kernel repetitions (p', e') with $\frac{n \cdot |e'| + \ell + k - 1}{n \cdot |p'|} = \frac{1}{k-1}$. Let (p', e') be a φ'-kernel repetition in w_{TM} with $|e'| \geq 4$

and $\frac{n\cdot|e'|+\ell+k-1}{n\cdot|p'|} = \frac{1}{k-1}$. By [28, Corollary 9], w_{TM} has a φ'-kernel repetition (p'', e'') with $|p'| = 2\cdot|p''|$ and $|e'| \leq 2\cdot|e''|$. Thus w_1 has a φ-kernel repetition $(h(p''), h(e'')L)$, and w_0 has a kernel repetition of exponent $\frac{n\cdot|e''|+\ell+k-1}{n\cdot|p''|} > \frac{n\cdot|e'|+\ell+k-1}{n\cdot|p'|} = \frac{1}{k-1}$. We have a contradiction with the fact that w_0 is a Dejean word. □

We apply the previous lemma on constructions for $8 \leq k \leq 38$ (Chapter 2), or $k \geq 24$ (Chapter 3) to show that $\mathrm{FRT}(k) = \mathrm{RT}(k)$ for every $k \geq 8$.

We conclude with the following open questions.

Conjecture 4.20. For every $k \geq 5$, there is a infinite Dejean word on k letters such that the only allowed limit repetitions have period $k-1$.

Let $\mathrm{LR}(k)$, for $k \geq 3$, be the minimal number of limit repetitions over all Dejean words on k letters. Similarly let $\mathrm{LR}'(k)$, for $k \geq 5$, be the minimal number of limit repetitions over all Dejean words on k letters such that every limit repetition has period $k-1$. By the results of the present article, $\mathrm{LR}(k)$ is defined for every $k \geq 3$, and $\mathrm{LR}'(k)$ is defined if Conjecture 4.20 is true. We know that $\mathrm{LR}(3) = 2$ [39], $\mathrm{LR}(4) = 2$, $45 \leq \mathrm{LR}(5) \leq 60$ and $46 \leq \mathrm{LR}'(5) \leq 60$. It may be difficult to find the exact value of $\mathrm{LR}(k)$ or $\mathrm{LR}'(k)$ for any $k \geq 5$, but we can ask the following question.

Question 4.21. Find a lower or an upper bound for $\mathrm{LR}(k)$ or $\mathrm{LR}'(k)$, $k \geq 5$.

Conjecture 4.20 implies that $\mathrm{LR}(k) \leq \mathrm{LR}'(k) \leq k!$. On the other hand, limit repetitions cannot be avoided when $k \geq 5$ since every 0 in the Pansiot code leads to an occurrence of a limit repetition of period $k-1$. Thus $0 < \mathrm{LR}(k) \leq \mathrm{LR}'(k)$.

Chapter 5

Others generalizations and open problems

This last chapter closes the first part dedicated to Dejean's conjecture. We present results and open questions about other generalizations of the problem.

5.1 Dejean words and letter frequency

Several authors studied the minimal (resp. maximal) frequency of a letter in a power-free language.

Let $\rho(x)$ (resp. $\rho^+(x)$) be the minimal frequency of a letter in a x-free (resp. x^+-free) word. The function ρ is introduced by Kolpakov, Kucherov and Tarannikov in [130], and studied in [122, 19]. The values of $\rho(x)$ are nearly known (see [19]). For example, one has $\rho(2^+) = \rho(7/3) = 1/2$, and $\rho(7/3^+) = 327/703 = 0.4651493598\ldots$. A related question is to find the possible frequencies of a letter in a square-free ternary word.

Theorem 5.1. *[106, 122]*

- *The minimal density of a letter in an infinite ternary square-free word is* $\frac{883}{3215} = 0.27465007\ldots$.
- *The maximal density of a letter in an infinite ternary square-free word is* $\frac{255}{653} = 0.39050535\ldots$.

Ochem's conjecture rises the question of minimal and maximal frequencies of a letter in Dejean words. The proof of this conjecture (Chapter 3) gives the answer for alphabets of size at least 5 for minimum frequency and size 6 for maximum frequency. But the question is still open for alphabets of size 3, 4 and 5.

Question 5.2. What is the minimal (resp. maximal) frequency of a letter in a Dejean word on k letters, for $k \in \{3, 4\}$ (resp. $k \in \{3, 4, 5\}$) ?

Ochem showed that the maximum frequency of a letter in a Dejean word on 5 letters is less than $\frac{103}{440} = 0.23409090\ldots$ [122].

5.2 Generalized repetition threshold

The generalized repetition threshold has been introduced by Ilie, Ochem and Shallit in [90].

A word is (α, ℓ)-free (resp. (α^+, ℓ)-free) if it does not contain any repetition of period at least ℓ and exponent at least α (resp. greater than α). The *generalized repetition threshold* $\mathrm{RT}(k, \ell)$ is the smallest real α such that there exists an infinite (α^+, ℓ)-free word on k-letters. The case $\ell = 1$ corresponds to Dejean's repetition threshold.

The general behaviour of $\mathrm{RT}(k, \ell)$ is known when ℓ tends to infinity [75], but the exact values are still unknown or conjectured, except for few cases. One has $\mathrm{RT}(2, 2) = 2$ (since one cannot avoid squares of period at least 2 on binary words), $\mathrm{RT}(2, 3) = \frac{8}{5}$, $\mathrm{RT}(2, 4) = \frac{3}{2}$, $\mathrm{RT}(2, 5) = \frac{7}{5}$, $\mathrm{RT}(2, 6) = \frac{4}{3}$, $\mathrm{RT}(3, 2) = \frac{3}{2}$ and $\mathrm{RT}(3, 3) = \frac{4}{3}$ [90]. Moreover, computations support the following intriguing conjecture.

Conjecture 5.3 ([90])**.** For $\ell \geq 2$, $\mathrm{RT}(3, \ell) = 1 + \frac{1}{\ell}$ and $\mathrm{RT}(4, \ell) = 1 + \frac{1}{\ell+2}$.

The following unpublished theorem gives a partial answer to the previous conjecture (joint work with Roman Kolpakov).

Theorem 5.4. *For all* $\ell \geq 11$, $\mathrm{RT}(3, \ell) \geq 1 + \frac{1}{\ell}$.

Proof. Suppose that $\mathrm{RT}(3, \ell) < 1 + \frac{1}{\ell}$ for a $\ell > 0$. Then for every $n > 0$ and $0 < m \leq \ell$, one can find n ternary words $u_1, u_2, \ldots, u_n$ of size m with the following property: for every $e > 0$ and $1 \leq i < j \leq n$ and $1 \leq x, y \leq m + 1 - e$ such that $u_i[x : x + e - 1] = u_j[y : y + e - 1]$, one has:

- $e < j - i + 1$ if $x < y$
- $e < j - i$ if either $x = y$ or $x > y$ and $i + 1 < j$.

To show this, it suffice to take $u_1 = w[1 : m], u_2 = [1 + \ell : m + \ell], ..., u_n = w[1 + (n - 1)\ell : m + (n - 1)\ell]$, where w is an infinite $(1 + 1/\ell, \ell)$-free ternary word.

A backtracking algorithm shows that such $u_1, u_2, \ldots, u_n$ do not exist for $n = 10$ and $m = 11$. Thus $\mathrm{RT}(3, \ell) \geq 1 + \frac{1}{\ell}$. □

It may be proved that $\mathrm{RT}(4, \ell) \geq 1 + \frac{1}{\ell+2}$ using a generalization of the previous technique, with an additional connectivity constraint on the Rauzy graph. Upper bounds of Conjecture 5.3 may be proved constructively. Thus, it is reasonable to think that Conjecture 5.3 can be solved.

5.3 Growth rate of Dejean words

The *growth rate* of a language L on an alphabet $\mathcal{A}$ is $\limsup_{n \to \infty} |L \cap \mathcal{A}^n|^{\frac{1}{n}}$. Some authors prefer the terminology of *entropy*, which is, in the case of languages on words, the logarithm of the growth rate. The growth rate of overlaps-free binary words is 1, since there are only polynomially many such words and the growth rate of cube-free binary words is between 1.45697 and 1.4576 [108, 69]. The growth rate of ternary square-free words is between 1.30173 and 1.30179 [108, 121].

We known that there are exponentially many Dejean words on k letters, for $k \in \{3, 4\}$ [120], $k \in \{5, \ldots, 10\}$ [99], and every odd $k \in \{7, \ldots, 101\}$ [144].

Conjecture 5.5. There are exponentially many Dejean words on k letters, for every $k \geq 3$.

More specifically, let g_k be the growth rate of Dejean words on a k-letter alphabet. One has bounds for g_k for some small k.

Theorem 5.6 ([120, 15])**.**

- $1.245 \leq g_3$
- $1 < g_4$
- $1.153811 \leq g_5 \leq 1.157895$
- $1.223437 \leq g_6 \leq 1.224695$
- $1.236409 \leq g_7 \leq 1.236899$

- $1.234725 \leq g_8 \leq 1.234843$
- $1.246659 \leq g_9 \leq 1.246678$
- $1.239287 \leq g_{10} \leq 1.239308$

For $k = 2$, $\mathrm{RT}(k) = 2$, and we are in the case of overlap-free binary words. As we already know, the growth rate is 1. More generally, there are polynomially many $\frac{7}{3}$-free binary words (their growth rate is then 1), and there are exponentially many $\frac{7}{3}^{+}$-free binary words [99]. The growth rate of $\frac{7}{3}^{+}$-free binary words is estimated at $1.2206448\ldots$ [139].

Moreover, computations strongly suggest that:

Conjecture 5.7 ([140])**.** $\lim_{k\to\infty} g_k = 1.242\ldots$.

Let $g = \limsup_{k\to\infty} g_k$. If one wants to compute an upper bound on g, one can focus on Pansiot code of Dejean words. Let w be the Pansiot code of Dejean word. As explained in Lemma 3.9, w forbids 00, 111 and $u\{0,1\}^{k-4}u$ for every $u \in \{0,1\}^4$. Thus, the growth rate g'_k of the language L_k of binary words avoiding 00, 111 and $\left\{u\{0,1\}^{k-4}u : u \in \{0,1\}^4\right\}$ is a upper bound for g_k, and $g \leq g'$ where $g' = \limsup_{k\to\infty} g'_k$. One can easily show that $\lim_{k\to\infty} g'_k$ exists, and is the growth rate of two-dimensional binary words avoiding 00, 111 and $\begin{smallmatrix} a & b & c & d \\ a & b & c & d \end{smallmatrix}$ for every $a, b, c, d \in \{0,1\}$.

Unfortunately, computing such growth rates on two-dimensional words is a difficult problem. One famous example is the computation of the growth rate φ_2 of the Fibonacci subshift on $\mathbb{Z}^2$, that is two-dimensional binary words avoiding 11 and $\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}$. The Fibonacci subshift on $\mathbb{Z}$ is well-known, and its growth rate is the golden mean. Computing the exact value of φ_2 is a long-standing open question (even if we have a good approximation: $\varphi_2 = 1.503048\ldots$) as well as in graph theory (the Fibonacci subshift corresponds then to stable sets in the grid) and in physics (the hard square model).

Part II

Abelian repetitions and generalizations

Chapter 6

Abelian power avoidability and Mäkelä's questions

In 1957 and 1961, Erdős asked two questions about the avoidability of squares in words, in the continuation of the works of Thue [71, 72]. Firstly, he asked if one can avoid arbitrarily long squares in binary words. Secondly, he asked if one can avoid abelian squares over a finite alphabet.

The subject of this second part is the avoidability of abelian repetitions and its generalizations.

6.1 Avoidability of abelian powers

Two words $u, v \in A^*$ are *abelian equivalent*, denoted $u \equiv_a v$, if for every $a \in A$, $|u|_a = |v|_a$, where $|u|_a$ is the number of occurrences of the letter a in the word u. A word u is an *abelian-n-th-power*, where $n \geq 2$, if $u = u_1 u_2 \dots u_n$ such that $u_i \equiv_a u_{i+1}$ for every $i \in \{1, \dots, n-1\}$. An *abelian square* (resp. *abelian cube*) is an abelian-2nd-power (resp. abelian-3rd-power). It is not difficult to see that every ternary word of size at least 8 has an abelian square.

Erdős [71, 72] raised the question whether abelian squares can be avoided in an infinite word on an alphabet of size 4. Evdokimov [73] showed that one can avoid them on an alphabet of size 25, which was later lowered to 5 by Pleasants [127]. Finally, Keränen [103] answered positively to Erdős's question in 1992, with the following construction, found with the help of a computer.

Theorem 6.1 (Keränen [103]). *Fixed points of the following* 85*-uniform morphism are abelian-square-free:*

$$\sigma_K : \begin{cases} a \to abcacdcbcdcadcdbdabacabadbabcbdbcbacbcdcacbabdabacadcbcdcacdbcbacbcdcacdcbdcdadbdcbca \\ b \to bcdbdadcdadbadacabcbdbcbacbcdcacdcbdcdadbdcbcabcbdbadcdadbdacdcbdcdadbdadcadabacadcdb \\ c \to cdacabadabacbabdbcdcacdcbdcdadbdadcadabacadcdbcdcacbadabacabdadcadabacabadbabcbdbadac \\ d \to dabdbcbabcbdcbcacdadbdadcadabacabadbabcbdbadacdadbdcbabcbdbcabadbabcbdbcbacbcdcacbabd \end{cases}$$

Moreover, Carpi showed that the number of abelian-square-free-words over 4 letters is exponential [48]. The best known lower bound on the growth rate, due to Keränen, is 1.02306 [105].

Besides that, Dekking answered to the question of the avoidability of abelian-n-th-powers, for $n \geq 3$.

Theorem 6.2 (Dekking [65]). *Fixed points of the following morphism are abelian-cube-free:*

$$\sigma_{D3} : \begin{cases} a \to & aabc \\ b \to & bbc \\ c \to & acc. \end{cases}$$

Theorem 6.3 (Dekking [65]). *The fixed point of the following morphism is abelian-4th-power free:*

$$\sigma_{D2} : \begin{cases} a \to & abb \\ b \to & aaab. \end{cases}$$

Moreover, the growth rate of abelian-cube-free words over 3 letters is at least $3^{1/19} = 1.059526\dots$ (see Section 7.3.2), and the growth rate of abelian-4th-power free binary words is at least $2^{1/16} = 1.044273\dots$ [60].

6.2 Decision algorithms

Dekking [65], and then Carpi [46] gave sufficient conditions for a morphism h to be abelian-n-th-power-free, that is for every word w, $h(w)$ is abelian-n-th-power-free if and only if w is abelian-n-th-power-free. If h is abelian-n-th-power-free then the fixed points of h are abelian-n-th-power-free, but the converse does not always hold. For example, $\sigma_4 : 0 \to 03, 1 \to 43, 3 \to 1, 4 \to 01$ is not abelian-cube-free (since, for example, $\sigma_4(1004) = 43030301$ contains the cube 303030), but the fixed point of σ_4 is cube-free (see Section 6.4).

The eigenvalues of a morphism $h : \Sigma^* \to \Sigma^*$ are the eigenvalues of the matrix M_h, such that for $x, y \in \Sigma$, $M_h[x, y] = |h(x)|_y$. Currie and Rampersad [59] gave an algorithm which decides, for a fixed integer n, if a fixed point of a morphism with no eigenvalue of absolute value at most 1 is abelian-n-th-power-free.

In order to attack the problems from Mäkelä (Section 6.3), we needed specifically to be able to decide on morphisms with some eigenvalues of absolute value less than 1. Using ideas of both [59] and [50], we showed the following.

Theorem 6.4 ([25])**.** *For any primitive morphism h with no eigenvalue of absolute value 1 it is possible to decide if the fixed points of h are abelian-k-th-power-free.*

Let σ_6 be the following morphism.

$$\sigma_6 : \begin{cases} a \to & ace, \qquad & b \to & adf \\ c \to & bdf, \qquad & d \to & bdc \\ e \to & afe, \qquad & f \to & bce. \end{cases}$$

Using decision algorithm from [25], we can show the following.

Theorem 6.5. *$\sigma_6^\omega(a)$ is abelian-square-free.*

In addition to being abelian-square-free, $\sigma_6^\omega(a)$ has an important property. The eigenvalues of σ_6 are 0 (with algebraic multiplicity 3), $-\sqrt{3}$, $\sqrt{3}$ and 3. Since σ_6 has only 3 eigenvalues (including multiplicities) of absolute value greater than 1, the Parikh vectors of the factors of $\sigma_6^\omega(a)$ are close to a subspace of $\mathbb{R}^6$ of dimension 3. This property is important to show Theorem 6.12 (one can avoid long abelian squares on a ternary alphabet), and Theorem 6.16 (one can avoid additive squares on $\mathbb{Z}^2$).

6.3 Avoidability of long abelian squares and questions of Mäkelä

Erdős also asked if it is possible to avoid arbitrarily long ordinary squares on binary words [71, 72]. Erdős thought that the answer was negative. Entringer, Jackson and Schatz showed the opposite: it is possible to construct an infinite binary word without squares of size 6 and more [70]. This result has been improved by Fraenkel and Simpson: it is possible to construct an infinite binary word with only 3 squares: 00, 11 and 1010, and this is the best we can do [77]. Perhaps the simplest construction of such a word is given by Badkobeh and Crochemore:

Theorem 6.6 ([38]). *Let* $\eta : 0 \to 01001110001101, 1 \to 0011, 2 \to 000111$. *Then* $\eta(w_{TTM})$ *contains only* 3 *squares:* 00*,* 11*, and* 1010*.*

In the same spirit Mäkelä asked the following two questions about the avoidability of long abelian cubes (resp. squares) on a binary (resp. ternary) alphabet:

Problem 6.7 (Mäkelä (see [104])). Can you avoid abelian-cubes of the form uvw where $|u| \geq 2$, over two letters ? - You can do this at least for words of length 250.

Problem 6.8 (Mäkelä (see [104])). Can you avoid abelian squares of the form uv where $|u| \geq 2$ over three letters ? - Computer experiments show that you can avoid these patterns at least in words of length 450.

We reformulated the questions of Mäkelä, and asked [24]:

Problem 6.9. Is there a $p \in \mathbb{N}$ such that one can avoid abelian squares of period at least p over three letters? If yes what is the smallest such p?

Problem 6.10. Is there a $p \in \mathbb{N}$ such that one can avoid abelian cubes of period at least p over two letters? If yes what is the smallest such p?

We showed that the answer to Question 6.7 is negative:

Theorem 6.11 ([24]). *There is no infinite word over a binary alphabet avoiding abelian cubes of period at least* 2*.*

This negative result is shown using an exhaustive search: one can show that there is an infinite word with the property if and only if there are infinitely many Lyndon words with the property, which is false.

On the other hand, we showed that a weakening version of the second question has a positive answer. Let σ_3 be the following morphism.

$$\sigma_3 : \begin{cases} a \to & \texttt{bbbaabaaac} \\ b \to & \texttt{bccacccbcc} \\ c \to & \texttt{ccccbbbcbc} \\ d \to & \texttt{ccccccccaa} \\ e \to & \texttt{bbbbbcabaa} \\ f \to & \texttt{aaaaaaabaa}. \end{cases}$$

Theorem 6.12 ([25]). $\sigma_3(\sigma_6^{\omega}(a))$ *does not contain any square of period more than* 5*.*

Thus we know that $2 \leq p \leq 6$ in Problem 6.8, and $p \geq 3$ in Problem 6.7. Theorem 6.12 is shown using a decision algorithm which is a simple extension of the algorithm of Theorem 6.4.

The technique we use in [25] to prove Theorem 6.12 may be used to prove Question 6.8 or to give a bound on p in Problem 6.10. The difficulty is to handle the combinatorial explosion when we search for a morphism (as σ_3 in Theorem 6.12).

6.4 Additive powers

Mäkelä's problems turned out to be close to problems asked by Justin, Pirillo and Varricchio. Let $k \geq 2$ be an integer and $(G, +)$ a group. An *additive k-th power* is a non-empty word $w_1 \dots w_k$ over $\mathcal{A} \subseteq G$ such that all for every $i \in \{2, \dots, k\}$, $|w_i| = |w_1|$ and $\sum w_i = \sum w_1$ (where $\sum v = \sum_{i=1}^{|v|} v[i]$). A group $(G, +)$ is *k-uniformly repetitive* if every infinite word over a finite subset of G contains an additive k-th power as a factor.

Problem 6.13 ([94, 126]). Can we avoid two consecutive blocks of the same size and the same sum over a finite subset of $\mathbb{Z}$? In other words, is $\mathbb{Z}$ non uniformly 2-repetitive ?

The answer of Problem 6.13 is likely negative, but two weakening of this problem are true.

Theorem 6.14 (Cassaigne *et al.*, 2014 [50]). *The fixed point of σ_4 is additive-cube free, with:*

$$\sigma_4 : \begin{cases} 0 \to & 03 \\ 1 \to & 43 \\ 3 \to & 1 \\ 4 \to & 01 \end{cases}$$

This implies that $\mathbb{Z}$ is not uniformly 3-repetitive.

We will see in Section 7.3 several ternary alphabets on which one can avoid additive cubes. It seems that it is easy to avoid additive cubes on the alphabet $\{0, i, j\}$, with $0 < i < j$, $j \geq 6$ and i co-prime with j, and this is proved for every $6 \leq j \leq 9$. Moreover, additives cubes are avoidable over $\{0, 1, 5\}$ (see Section 7.3.2). In [25, 135], one show that one can avoid additive cubes on $\{0, 1, 2, 4\}$, $\{0, 2, 3, 5\}$ and $\{0, 2, 3, 6\}$. This leaves open the following question, for which it seems difficult to find a construction.

Question 6.15. Is there infinite additive-cube-free words on the following alphabets : $\{0, 1, 2, 3\}$, $\{0, 1, 4\}$ and $\{0, 2, 5\}$?

Let Φ the morphism such that:

$$\Phi : \begin{cases} a \to & (0,0) & \qquad b \to & (1,1) \\ c \to & (2,1) & \qquad d \to & (0,1) \\ e \to & (2,0) & \qquad f \to & (1,0). \end{cases}$$

Theorem 6.16 ([25]). $\Phi(\sigma_6^\omega(a))$ *does not contain two consecutive blocks of the same size and the same sum. In other words, $\mathbb{Z}^2$ is not uniformly 2-repetitive.*

A simple extension of algorithm of Theorem 6.4 is able to prove Theorem 6.14 and Theorem 6.16, and it is not a coincidence that σ_6 is the base of constructions of Theorem 6.12 and Theorem 6.16.

Again, it would be interesting to find smaller alphabets $\Sigma \subseteq \mathbb{N}^2$ on which one can avoid additive squares. Surely, a method similar to the one presented in Section 7.3 would give a positive answer for some alphabets of size 5. Nevertheless, it would be difficult to find alphabets of size 4, since this result would imply a new construction of a square-free word on 4 letters.

6.5 Parameterized generalizations

Recently, two parameterized variations of the abelian equivalence have been introduced: the *k-abelian equivalence*, introduced by Karhumäki *et al.* [88, 98, 97], and the *k-binomial equivalence*, introduced by Rigo and Salimov [133]. These two notions bring a gap between the abelian equivalence (which is the 1-abelian equivalence and the 1-binomial equivalence) and the usual equality between words (which can be viewed as the ∞-abelian equivalence, or the ∞-binomial equivalence). Moreover, these two notions are not comparable, except for $k = 1$.

The next two chapters are devoted to the avoidability of powers with respect to the k-abelian-equivalence (Chapter 7) and the k-binomial-equivalence (Chapter 8).

Chapter 7

Avoiding k-abelian powers

Carpi gave a set of conditions which imply that a morphism h is abelian-power-free, that is $h(w)$ is abelian-n-th-power-free if and only if w is abelian-n-th-power-free [46]. Moreover, this set is conjectured to be a characterization of abelian-power-free morphisms. In this chapter, we adapt this set of conditions to k-abelian-repetitions: we give sufficient conditions for a morphism h to be k-abelian-n-th-power-free, that is, for every abelian-n-th-power-free word w, $h(w)$ is k-abelian-n-th-power-free. In a very similar way, we give sufficient conditions for a morphism h to be additive-n-th-power-free, that is, for every additive-n-th-power-free word w, $h(w)$ is also additive-n-th-power-free.

Using these results, we prove that 2-abelian-cubes are avoidable over a binary alphabet and that 3-abelian-squares are avoidable over a ternary alphabet, answering positively to two questions of Karhumäki *et al.*. We also show the existence of infinite additive-cube-free words on several ternary alphabets.

Additionally, all our constructions show that the number of such words grows exponentially. As a corollary, we get a new lower bound of $3^{1/19} = 1.059526\ldots$ for the growth rate of abelian-cube-free words.

This chapter is based on the paper [23].

7.1 Introduction

We are here interested in two variations of the problem of abelian power avoidability. The first one is the k-abelian-equivalence introduced by Karhumäki *et al.* [88, 98, 97]. Let $k \geq 1$. Two words u and v ($u, v \in A^*$) are ***k*-abelian-equivalents**, denoted $u \equiv_{a,k} v$, if for every $w \in A^*$ with $|w| \leq k$, $|u|_w = |v|_w$. A word u is a ***k*-abelian-*n*-th-power**, $n \geq 2$, if $u = u_1 u_2 \ldots u_n$ such that $u_i \equiv_{a,k} u_{i+1}$ for every $i \in \{1, \ldots, n-1\}$. A ***k*-abelian-square** (resp. ***k*-abelian-cube**) is a k-abelian-2nd-power (resp. k-abelian-3rd-power). This notion is between the abelian equivalence (which is the 1-abelian-equivalence) and the usual equality between words (which can be viewed as the ∞-abelian-equivalence). Since cubes are avoidable in the binary alphabet (*e.g.* in the Prouhet-Thue-Morse word), but are not avoidable in the abelian sense, it is natural to ask for the smallest k for which k-abelian-cubes are avoidable on a binary alphabet. In [88] authors showed that $k \leq 8$, and in [116] that $k \leq 5$. Finally, in [115], Mercaş and Saarela showed that $k \leq 3$. The same question can be asked for k-abelian-squares on a ternary alphabet: 2-abelian-squares cannot be avoided [89], but Huova showed that 64-abelian-squares can be avoided [87].

In Section 7.2, we give sufficient conditions for a morphism $h : A^* \to B^*$ to be k-abelian-n-th-power-free (for a fixed $n \geq 2$ and $k \geq 1$), that is for every abelian-n-th-power-free word $w \in A^*$, $h(w)$ is k-abelian-n-th-power-free. Then we give morphisms which respect the conditions, in order to construct 2-abelian-cube-free binary words and 3-abelian-square-free ternary words. This gives the answer to the two previous questions, and prove on the same time that the number of such words grows exponentially, as abelian-square-free on four letters [48], and abelian-cube-free ternary words ([34], see also Section 7.3).

The second notion is the additive-cube-avoidability. A word $w \in \mathbb{N}^*$ is an *additive cube* if $w = pqr$, where p, q and r are non-empty-word such that $|p| = |q| = |r|$ and $\sum(p) = \sum(q) = \sum(r)$. A word is *additive-cube-free* if it has no factor which is an additive cube. Clearly, such words are also abelian-cube-free. Recently Cassaigne *et al.* [50] showed that one can construct an infinite additive-cube-free word on the alphabet $\{0, 1, 3, 4\}$. The question if there exists an infinite additive-square-free word on a finite alphabet is still open.

In Section 7.3 we give sufficient conditions for a substitution $h : A^* \to 2^{B^*}$, $A, B \subseteq \mathbb{N}$, to be additive-cube-free. We present substitutions from the alphabet $\{0, 1, 3, 4\}$ to several ternary alphabets which respects these conditions. Moreover, the presented constructions show directly that the number of additive-cube-free words on these ternary alphabets grows exponentially. The lower bound of $3^{1/19} = 1.059526\ldots$ that we obtain for the growth rate for the alphabet $\{0, 1, 8\}$ is also a new lower bound for the number of abelian-cube-free words on a ternary alphabet.

7.2 k-abelian-n-th-power-free morphisms

7.2.1 Preliminaries

Let $|u|_w$ denote the number of occurrences of the factor w in u. The *Parikh vector* of a word $u \in A^*$, where $A = \{a_1, \ldots, a_k\}$, is $\Psi(u) = (|u|_{a_1}, \ldots, |u|_{a_k})$. For a set $S \subseteq A^*$, $\Psi_S(u)$ is the vector indexed by S such that $\Psi_S(u)[w] = |u|_w$ for every $w \in S$. When the alphabet is clear in the context, we let $\Psi_k(u)$ be $\Psi_{A^k}(u)$, for $k \geq 1$.

Let $\mathrm{Pref}(u)$ be the set of prefixes of u, and $\mathrm{Suf}(u)$ be its set of suffixes. For $k \geq 0$, let $\mathrm{pref}_k(u)$ (resp. $\mathrm{suf}_k(u)$) be the prefix (resp. suffix) of u of size k.

There are several equivalent definitions for k-abelian-equivalence (see [97]). Two words u and v of size at most $k-1$ are k-abelian-equivalent if and only if they are equal. Otherwise, the following conditions are equivalent:

- u and v are k-abelian-equivalent (*i.e.* $u \equiv_{a,k} v$).
- For every $w \in A^*$ with $|w| \leq k$, $|u|_w = |v|_w$.
- For every $w \in A^k$, $|u|_w = |v|_w$, $\mathrm{pref}_{k-1}(u) = \mathrm{pref}_{k-1}(v)$ and $\mathrm{suf}_{k-1}(u) = \mathrm{suf}_{k-1}(v)$.
- For every $w \in A^k$, $|u|_w = |v|_w$, and $\mathrm{pref}_{k-1}(u) = \mathrm{pref}_{k-1}(v)$.

Given $k \geq 1$ and $n \geq 2$, a (possibly infinite) word w is *k-abelian-n-th-power-free* if no non-empty factor in w is a k-abelian-n-th-power. A word is *k-abelian-square-free* (resp. *k-abelian-cube-free*) if it is k-abelian-2nd-power-free (resp. k-abelian-3rd-power-free).

A morphism $h : A^* \to B^*$ is *k-abelian-n-th-power-free* if for every abelian-n-th-power-free word $u \in A^*$, $h(u)$ is k-abelian-n-th-power-free. Note that u has to be abelian-n-th-power-free, not only k-abelian-n-th-power-free; we explain in Section 7.2.4 why we use this weaker notion. A morphism $h : A^* \to B^*$ is *k-abelian-square-free* (resp. *k-abelian-cube-free*) if it is k-abelian-2nd-power-free (resp. k-abelian-3rd-power-free).

7.2.2 Testing k-abelian-n-th-power-freeness

In [46], Carpi gave a set of conditions which assure that a given morphism is abelian-n-th-power-free. We give in the following theorem a set of similar conditions which assure that a given morphism is k-abelian-n-th-power-free.

Theorem 7.1. *We fix $k \geq 1$ and $n \geq 2$, and two alphabets A and B. Let $h : A^* \to B^*$ be a morphism. Suppose that:*

(i) For every abelian-n-th-power-free word $w \in A^*$ with $|w| \leq 2$ or $|h(w[2 : |w|-1])| \leq (k-2)n - 2$, $h(w)$ is k-abelian-n-th-power-free.

(ii) There are $p, s \in B^{k-1}$ such that for every $a \in A$, $p = \mathrm{pref}_{k-1}(h(a)p)$ and $s = \mathrm{suf}_{k-1}(sh(a))$.

(iii) The matrix N indexed by $B^k \times A$, with $N[w,x] = |h(x)p|_w$, has rank $|A|$.

(iv) Let $S \subseteq B^k$, with $|S| = |A|$, such that the matrix M indexed by $S \times A$, with $M[w,x] = |h(x)p|_w$, is invertible. Let

$$\Psi_S(v,u) = \Psi_S(vp) + \Psi_S(su) - \Psi_S(sp)$$

and $\Psi_k(v,u) = \Psi_{B^k}(v,u)$. For every $a_i \in A$ and $u_i, v_i \in A^*$ with $u_i v_i = h(a_i)$; $0 \leq i \leq n$; such that:

(P) $|\{\mathrm{pref}_{k-1}(v_i p) : 0 \leq i < n\}| = 1$,

(I) $M^{-1}(\Psi_S(v_{i-1}, u_i) - \Psi_S(v_i, u_{i+1}))$ is an integer vector, for every $1 \leq i < n$,

(C) $\Psi_k(v_{i-1}, u_i) - \Psi_k(v_i, u_{i+1}) \in \mathrm{im}(N)$ for every $1 \leq i < n$,

there is a $(\alpha_0, \ldots, \alpha_n) \in \{0,1\}^{n+1}$ such that for every $1 \leq i < n$:

$$\begin{aligned} M^{-1}\Psi_S(v_{i-1}, u_i) - (1-\alpha_{i-1})\Psi(a_{i-1}) - \alpha_i \Psi(a_i) \\ = M^{-1}\Psi_S(v_i, u_{i+1}) - (1-\alpha_i)\Psi(a_i) - \alpha_{i+1}\Psi(a_{i+1}). \end{aligned} \tag{7.1}$$

Then h is k-abelian-n-th-power-free.

Proof. Suppose that $h(w)$ has a k-abelian-n-th-power $q_1 \ldots q_n$. Let q_0 and q_{n+1} be such that $h(w) = q_0 q_1 \ldots q_n q_{n+1}$. By condition (i), if $|q_i| < k-1$, then w has an abelian-n-th-power. So we have $|q_i| \geq k-1$.

There are, for every $0 \leq i \leq n$, $a_i \in A$, $u_i \in \mathrm{Pref}(h(a_i))$ and $r_i \in A^*$ such that, for every $0 \leq i \leq n$, $r_0 \ldots r_i a_i \in \mathrm{Pref}(w)$ and $q_0 \ldots q_i = h(r_0 \ldots r_i)u_i$. Note that, for a $1 \leq i \leq n$, r_i can be empty, but a_i is always the first letter of $r_{i+1}a_{i+1}$. Let v_i be such that $u_i v_i = h(a_i)$ for every $0 \leq i \leq n$. By condition (i), one can suppose w.l.o.g. that $|r_1 \ldots r_n a_n| \geq 3$.

By condition (ii), for every $1 \leq i \leq n$, $\mathrm{pref}_{k-1}(q_i) = \mathrm{pref}_{k-1}(v_{i-1}p)$. Since $q_1 \ldots q_n$ is a k-abelian-n-th-power, we have condition (P).

Claim 7.2. Let $r \in A^*$ and $u, v \in B^*$. Then:

- $N\Psi(r) = \Psi_k(h(r)p) = \Psi_k(sh(r)) = \Psi_k(sh(r)p) - \Psi_k(sp)$,
- $\Psi_k(vh(r)p) = \Psi_k(vp) + N\Psi(r)$,
- $\Psi_k(sh(r)u) = \Psi_k(su) + N\Psi(r)$.

Proof. If $\mathrm{pref}_{k-1}(u) = p$, then $\Psi_k(vu) = \Psi_k(vp) + \Psi_k(u)$. Similarly, if $\mathrm{suf}_{k-1}(v) = s$, then $\Psi_k(vu) = \Psi_k(v) + \Psi_k(su)$. All the equality follow from the previous facts, and the definition of N. □

Claim 7.3. For every $1 \le i \le n$:

$$\Psi_k(q_i) = N(\Psi(r_i) - \Psi(a_{i-1})) + \Psi_k(v_{i-1}, u_i). \tag{7.2}$$

Proof. By double counting, we have :

$$\Psi_k(q_i) + \Psi_k(sh(r_i a_i)p) = \Psi_k(sh(r_i)u_i) + \Psi_k(v_{i-1}h(a_{i-1}^{-1} r_i a_i)p).$$

By Claim 7.2:

$$\Psi_k(q_i) + N\Psi(r_i a_i) + \Psi_k(sp) = \\ \Psi_k(su_i) + N\Psi(r_i) + \Psi_k(v_{i-1}p) + N\Psi(a_{i-1}^{-1} r_i a_i).$$

Thus: $\Psi_k(q_i) = \Psi_k(v_{i-1}, u_i) + N(\Psi(r_i) - \Psi(a_{i-1}))$. □

Since $\Psi_k(q_i) = \Psi_k(q_{i+1})$ for every $1 \le i < n$, we have the condition (C). Now we have directly $\Psi_S(q_i) = M(\Psi(r_i) - \Psi(a_{i-1})) + \Psi_S(v_{i-1}, u_i)$. Since $\Psi_S(q_i) = \Psi_S(q_{i+1})$:

$$M^{-1}(\Psi_S(v_{i-1}, u_i) - \Psi_S(v_i, u_{i+1})) = \Psi(r_{i+1}) - \Psi(a_i) - \Psi(r_i) + \Psi(a_{i-1}).$$

The right part is an integer vector, so we have condition (I). Thus, by condition (iv), there is $(\alpha_0, \ldots, \alpha_n) \in \{0,1\}^{n+1}$ such that (7.1) is fulfilled.

Equation (7.1) together with equations (7.2) give:

$$- \Psi(r_i) + \Psi(a_{i-1}) - (1 - \alpha_{i-1})\Psi(a_{i-1}) - \alpha_i \Psi(a_i) \\ = -\Psi(r_{i+1}) + \Psi(a_i) - (1 - \alpha_i)\Psi(a_i) - \alpha_{i+1}\Psi(a_{i+1})$$

that is:

$$\Psi(r_i) - \alpha_{i-1}\Psi(a_{i-1}) + \alpha_i \Psi(a_i) = \Psi(r_{i+1}) - \alpha_i \Psi(a_i) + \alpha_{i+1}\Psi(a_{i+1}). \tag{7.3}$$

In equation (7.3), either the left of the right part is a non-negative vector. Since equation (7.3) is fulfilled for every $1 \le i < n$, $\Psi(r_i) - \alpha_{i-1}\Psi(a_{i-1}) + \alpha_i \Psi(a_i)$ is a non negative vector for every $1 \le i \le n$. Let $r'_i = a_{i-1}^{-\alpha_{i-1}} r_i a_i^{\alpha_i}$; $1 \le i \le n$. Since a_i is the first letter of $r_i a_{i+1}$, and $\Psi(r'_i) = \Psi(r_i) - \alpha_{i-1}\Psi(a_{i-1}) + \alpha_i \Psi(a_i)$ is a non-negative vector, r'_i is well defined in B^*. In one hand $r'_1 \ldots r'_n$ is a factor of w, and is non empty since $|r'_1 \ldots r'_n| \ge |r_1 \ldots r_n a_n| - 2$. On the other hand $\Psi(r'_i) = \Psi(r'_{i+1})$ (by equation 7.3), for every $1 \le i < n$. Thus w has an abelian-n-th-power $r'_1 \ldots r'_n$. □

We introduce $\Psi_S(v, u)$ in order to handle pairs (v, u) such that $|vu| < k - 1$ (otherwise we have $\Psi_k(v, u) = \Psi_k(vu)$). Theorem 7.1 gives a set of sufficient conditions, but are still far from a characterization, as Carpi partially done for abelian-n-th-power-free morphisms [46]. The key point is the condition (ii). One mention that we can save up the suffix condition in (ii) by carefully handing the cases where u_i or v_i has size less than k. However we still need either the prefix (or the suffix) condition in order to properly define N.

7.2.3 2-abelian-cube-free and 3-abelian-square-free morphisms

Morphisms h_2 and h'_2 respect the conditions of Theorem 7.1 for $k = 2$ and $n = 3$, *i.e.* are 2-abelian-cube-free, while morphisms h_3 and h'_3 respect the conditions for $k = 3$ and $n = 2$, *i.e.* are 3-abelian-square-free. The checks were done by computer, and took few seconds. Thus the infinite word $h_2(u)$ (resp. $h'_2(u)$) where u is an infinite abelian-cube-free word (for example a fixed point of Dekking's morphism $\mu : 0 \to 0012,\ 1 \to 112,\ 2 \to 022$ [65]) is a 2-abelian-cube-free binary word. Similarly, $h_3(v)$ (resp. $h'_3(v)$), where v is an infinite abelian-square-free word on an alphabet of size 4 (for example, a fixed point of Keränen's morphism g_{85} [103]), is an infinite 3-abelian-square-free ternary word.

Over all the 2-abelian-cube-free morphisms we found, h_2 is the smallest uniform morphism, while h'_2 is the one which minimize $|h(012)|$. If we are only interested in 2-abelian-cube-free infinite word, one can find simpler construction. The morphism $h_d \circ \mu$ is 2-abelian-cube-free so $h_d(\mu^\infty(0))$ is 2-abelian-cube-free.

We also claim that $h'_d(\mu^\infty(0))$ is 2-abelian-cube-free. One can modify the decision procedure of Theorem 7.1 to compute the set of "patterns" that u has to avoid to ensure that $h(u)$ is k-abelian-n-th-power-free. This notion of patterns was used by Carpi [48, 47] to prove that a substitution is abelian-square free, or by Keränen [104] to prove that a fixed point of g_{98} is abelian-square free, even though g_{98} is not abelian-square free. This was also used, under the name of *template*, by Aberkane *et al.* [34] to show the exponential growth rate of abelian-cube-free ternary words, and by Currie and Rampersad [59] for an algorithm which decide if a fixed point of a morphism is abelian-n-th-power-free. More recently, Mercaş and Saarela [116, 115] used this kind of patterns to show that a morphic word is k-abelian-cube-free.

Doing this, we are able to show that $h'_d \circ \mu^3(u)$ is 2-abelian-cube free if and only if u forbids factors of the form $F = \{pqr, 1p0q0r2 : \Psi(p) = \Psi(q) = \Psi(r)\} \cup \{0p1q0r2, 1p1q0r2 : \Psi(p1) = \Psi(q0) = \Psi(r0)\}$. Moreover, $\mu(u)$ forbids factors of the form F if and only if u forbids factors of the form F (in others words, μ is "F-free"). Thus $h'_d(\mu^\infty(0))$ is 2-abelian-cube-free, but for every $n \geq 0$, $h'_d \circ \mu^n$ is not 2-abelian-cube-free (*e.g.* for every $n \geq 0$, $h'_d(\mu^n(1002))$ has a 2-abelian-cube).

7.2.4 Final remarks and questions

We finally shortly explain why we use this weak notion of k-abelian-n-th-power-freeness for morphisms. On one hand, k-abelian-squares cannot be avoided by a pure morphic word on a ternary alphabet [114]. So there is no morphism $h : \{0,1,2\} \to \{0,1,2\}$ such that for every k-abelian-square-free word u, $h(u)$ is k-abelian-square-free, except trivial ones. One the other hand, suppose that there is a morphism $h : A^* \to B^*$, with $|A| > |B|$, such that for every 2-abelian-cube-free word $u \in A^*$, $h(u)$ is 2-abelian-cube-free. Without lost of generality, there is $\{a, b\} \subseteq A$, such that the first letter of $h(a)$ and $h(b)$ is the same. Then $babbababb$ is 2-abelian-cube-free, but $h(bab) \equiv_{a,2} h(abb)$ thus $h(babbababb)$ is an 2-abelian-cube. We have a contradiction, so such a morphism cannot exist. Nevertheless we cannot conclude directly when $|A| = |B|$ and the first and last letters of the images differ. More specifically, the following question is still open.

2-abelian-cube-free morphisms:

$$h_2 : \begin{cases} 0 \to 001001010010110010010100100110010011001011\,01011 \\ 1 \to 001001100100110011011001101100100110011011\,01011 \\ 2 \to 001101101011010010110101101001010010011011\,01011 \end{cases}$$

$$h'_2 : \begin{cases} 0 \to 00100101001100100101001001100100110011011 \\ 1 \to 010110110011011001100100110011011 \\ 2 \to 0101101001010010110011011 \end{cases}$$

3-abelian-square-free morphisms:

$$h_3 : \begin{cases} 0 \to 01020120210120102012 10212 \\ 1 \to 0102101201021201210120212 \\ 2 \to 0102101210212021020120212 \\ 3 \to 0121020120210201210120212 \end{cases}$$

$$h'_3 : \begin{cases} 0 \to 012010201202120121012010 21 \\ 1 \to 0120212012102120102 1 \\ 2 \to 0120210201021 \\ 3 \to 0121020121 \end{cases}$$

Morphisms such that $h(\mu^\infty(0))$ is 2-abelian-cube-free:

$$h_d : \begin{cases} 0 \to 00100110011011001100100110010 0101 \\ 1 \to 00101101011010010100100110010 0101 \\ 2 \to 00101101011011001100100110101 1011 \end{cases}$$

$$h'_d : \begin{cases} 0 \to 0101101001011 \\ 1 \to 010110110011011001100100110011011 \\ 2 \to 0010010100100110010011001 1011 \end{cases}$$

Table 7.1 – Morphisms for k-abelian-n-th-power-free words.

Question 7.4. It there a pure morphic binary word which avoids 2-abelian-cubes ?

Following Erdős's and Mäkelä's questions, one can also ask for which $k \geq 2$ long k-abelian squares are avoidable over a binary alphabet. We showed that the answer is yes for every $k \geq 2$ [24, 25]. More specifically, let $g(k)$ be the least integer such that there exists an infinite binary word with only $g(k)$ distinct k-abelian squares. We showed the following.

Theorem 7.5. *[24, 25] Then $g(1) = \infty$, $5 \leq g(2) \leq 734$, $g(4) = g(3) = 4$, and $g(k) = 3$ for every $k \geq 5$.*

The determination of the exact value $g(2)$ is still open, and may have the same order of difficulty as Problem 6.9 and Problem 6.10.

7.3 Ternary words avoiding additive cubes

7.3.1 Testing additive-n-th-power-freeness

Let Σ be the morphism from the free monoid on the alphabet $\mathbb{N}$ and the additive group $(\mathbb{Z}, +)$ such that $\Sigma(x) = x$ for every $x \in \mathbb{N}$. A word $w \in \mathbb{N}^*$ is an *additive-n-th-power*, with $n \geq 2$, if $w = p_1 \ldots p_n$, such that for every $1 \leq i < n$, $|p_i| = |p_{i+1}|$ and $\Sigma(p_i) = \Sigma(p_{i+1})$. A word is an additive-cube (resp. additive-square) if it is an additive-3rd-power (additive-2nd-power). A (possibly infinite) word w is *additive-n-th-power-free* if no non-empty factor of w is an additive-n-th-power. Clearly, such words are also abelian-n-th-power-free. In [50], authors prove that the fixed point of the morphism $\sigma_4 : 0 \to 03, 1 \to 43, 3 \to 1, 4 \to 01$ is additive-cube-free.

A *substitution* is a morphism $s : A^* \to 2^{B^*}$ between the free monoid A^* and the power monoid of B^*, that is the monoid of subsets of B^*, with the operation $U \cdot V = \{uv : (u, v) \in U \times V\}$. A morphism $h : A^* \to B^*$ can be viewed as a substitution $s : A^* \to 2^{B^*}$ such that $s(w) = \{h(w)\}$. A substitution $s : A^* \to 2^{B^*}$, where $A, B \subseteq \mathbb{N}$, is *additive-$n$-th-power-free* if for every additive-n-th-power-free word $u \in A^*$, every $v \in s(u)$ is additive-n-th-power-free.

We give sufficient conditions for a substitution to be additive-n-th-power-free in the following theorem.

Theorem 7.6. *We fix $n \geq 2$ and $A, B \subseteq \mathbb{N}$. Let $s : A^* \to 2^{B^*}$ be a substitution. Suppose that:*

(i) For every additive-n-th-power-free word $w' \in A^*$ with $|w| \leq 2$, every $w \in s(w')$ is additive-n-th-power-free.

(ii) There is $(l, \gamma, \beta) \in \mathbb{N} \times \mathbb{Z} \times \mathbb{Z}$, with $\beta \neq 0$, such that for every $a \in A$ and $w \in s(a)$, we have $|w| = l$ and $\Sigma(w) = \gamma + a\beta$.

(iii) For every $a_i \in A$, $w_i \in s(a_i)$, and $u_i, v_i \in A^*$ with $u_i v_i = w_i$; $0 \leq i \leq n$; such that for every $1 \leq i < n$:

(L) $|v_{i-1} u_i| \equiv |v_i u_{i+1}| \pmod{l}$,

(M) $\Sigma(v_{i-1} u_i) \equiv \Sigma(v_i u_{i+1}) + x_i \gamma \pmod{\beta}$,

$$s_{015} : \begin{cases} 0 \to \{005015100100115010115, 005015100100115100115\} \\ 1 \to \{005015100100105055115, 050015100100105055115\} \\ 3 \to \{005015101155155055115, 050015101155155055115\} \\ 4 \to \{005015155055155055115, 050015155055155055115\} \end{cases}$$

$$s_{016} : \begin{cases} 0 \to \{0010116010100600101 6, 0010116010100600110 6\} \\ 1 \to \{00166060101006001016, 00166060101006001106\} \\ 3 \to \{00166166110160661106, 00166166110166061106\} \\ 4 \to \{00166166066160661106, 00166166066166061106\} \end{cases}$$

$$s_{017} : \begin{cases} 0 \to \{00170010011711001071, 00170010011711001701\} \\ 1 \to \{00170017707001001071, 00170017707001001701\} \\ 3 \to \{00170017711017177077, 01070017711017177077\} \\ 4 \to \{00170017707077177077, 01070017707077177077\} \end{cases}$$

$$s_{027} : \begin{cases} 0 \to \{0020720220220722007, 0020720220227022007\} \\ 1 \to \{7220720220220722007, 7220720220227022007\} \\ 3 \to \{7077200770720722007, 7077200770727022007\} \\ 4 \to \{7077272770720722007, 7077272770727022007\} \end{cases}$$

$$s_{037} : \begin{cases} 0 \to \{00300307303037707307, 00300307303037700737, 00300307303037707037\} \\ 1 \to \{00300300707737700737, 00300300707737707037, 00300300707737707307\} \\ 3 \to \{00337730337737700737, 00337730337737707037, 00337730337737707307\} \\ 4 \to \{00337737707737700737, 00337737707737707307, 00337737707737707037\} \end{cases}$$

$$s_{018} : \begin{cases} 0 \to \{0081001008011811011, 0081010080011811011, 0081001080011811011\} \\ 1 \to \{0081001008011818008, 0081010080011818008, 0081001080011818008\} \\ 3 \to \{0081018818808811811, 0081108818808811811, 0081810818808811811\} \\ 4 \to \{0081018818808808188, 0081108818808808188, 0081188018808808188\} \end{cases}$$

$$s_{038} : \begin{cases} 0 \to \{003800303830033833003, 003800308330033833003\} \\ 1 \to \{003800303830080038388, 003800308330080083838\} \\ 3 \to \{003808833833038838838, 003808838330338838838\} \\ 4 \to \{003808838388088388388, 083008838388088388388\} \end{cases}$$

$$s_{019} : \begin{cases} 0 \to \{0090110191001009, 0090110911001009\} \\ 1 \to \{0090119110110199, 0900119110110199\} \\ 3 \to \{0090190090099199, 0900190090099199\} \\ 4 \to \{0090119199099199, 0900119199099199\} \end{cases}$$

$$s_{029} : \begin{cases} 0 \to \{00290020020090022029, 00290020020090020229\} \\ 1 \to \{00290099220090022029, 00290099220090020229\} \\ 3 \to \{00220292299099299099, 22920220099099299099\} \\ 4 \to \{22920992299099299099, 22990292299099299099\} \end{cases}$$

$$s_{049} : \begin{cases} 0 \to \{00400400900499009, 00400400900949009\} \\ 1 \to \{00400449440099409, 00400449440499009\} \\ 3 \to \{00409909909499409, 00409909949099409\} \\ 4 \to \{44944909949499009, 44944909949909409\} \end{cases}$$

Table 7.2 – Additive-cube-free substitutions.

(where $x_i = (|v_{i-1}u_i| - |v_i u_{i+1}|)/l$ *for every* $1 \leq i < n$*)*
there is $(\alpha_0, \ldots, \alpha_n) \in \{0,1\}^{n+1}$ *such that for every* $1 \leq i < n$*:*

(a) $\alpha_i - \alpha_{i-1} = x_i + \alpha_{i+1} - \alpha_i$,

(b) $\Sigma(v_{i-1}u_i) + \beta[(\alpha_{i-1} - 1)a_{i-1} - \alpha_i a_i]$
$= \Sigma(v_i u_{i+1}) + \gamma x_i + \beta[(\alpha_i - 1)a_i - \alpha_{i+1}a_{i+1}]$.

Then s *is additive-*n*-th-power-free.*

Proof. Suppose that $w \in s(w')$ has an additive-n-th-power $q_1 \ldots q_n$. Let q_0 and q_{n+1} be such that $w = q_0 q_1 \ldots q_n q_{n+1}$.

For every $0 \leq i \leq n$, there is $a_i \in A$, $w_i \in s(a_i)$, $u_i \in \mathrm{Pref}(w_i)$, and $r_i \in A^*$ such that $r_0 \ldots r_i a_i \in \mathrm{Pref}(w')$ and $q_0 \ldots q_i \in s(r_0 \ldots r_i) \cdot \{u_i\}$. Let v_i be such that $u_i v_i = w_i$ for every $0 \leq i \leq n$. By condition (i), one can suppose w.l.o.g. that $|r_1 \ldots r_n a_n| \geq 3$.

By condition (ii), for every $p \in s(p')$, we have $\Sigma(p) = \gamma|p'| + \beta\Sigma(p')$.

For every $1 \leq i \leq n$, we have $u_{i-1}q_i \in s(r_i) \cdot \{u_i\}$. Thus, by condition (ii), and by the fact that $u_{i-1}v_{i-1} = w_{i-1}$ we have:

$$|q_i| = |v_{i-1}u_i| + l(|r_i| - 1) \tag{7.4}$$

and

$$\Sigma(q_i) = \gamma(|r_i| - 1) + \beta(\Sigma(r_i) - a_{i-1}) + \Sigma(v_{i-1}u_i). \tag{7.5}$$

By equation (7.4) and by the fact that for every $1 \leq i < n$, $|q_i| = |q_{i-1}|$, we have the condition (L), and:

$$|r_{i+1}| - |r_i| = (|v_{i-1}u_i| - |v_i u_{i+1}|)/l = x_i.$$

Since for every $1 \leq i < n$, $\Sigma(q_i) = \Sigma(q_{i-1})$, we have:

$$\begin{aligned}\gamma(|r_i| - 1) + \beta(\Sigma(r_i) - a_{i-1}) + \Sigma(v_{i-1}u_i)& \\ = \gamma(|r_{i+1}| - 1) + \beta(\Sigma(r_{i+1}) - a_i) + \Sigma(v_i u_{i+1}).&\end{aligned} \tag{7.6}$$

Thus

$$\Sigma(v_{i-1}u_i) = \Sigma(v_i u_{i+1}) + \gamma x_i + \beta(\Sigma(r_{i+1}) - a_i - \Sigma(r_i) + a_{i-1}),$$

and equation (M) is fulfilled.

So, by condition (iii), there is $(\alpha_0, \ldots, \alpha_n) \in \{0,1\}^{n+1}$ such that (a) and (b) are fulfilled.

By equation (a), we have, for every $1 \leq i < n$;

$$|r_i| + \alpha_i - \alpha_{i-1} = |r_{i-1}| + \alpha_{i+1} - \alpha_i \tag{7.7}$$

If r_i is empty, $a_i = a_{i+1}$ otherwise the first letter of r_i is a_i. In equation (7.7), the right side or the left side must be non-negative. Thus for every $1 \leq i \leq n$, $|r_i| + \alpha_i - \alpha_{i-1} \geq 0$, and $r'_i = a_{i-1}^{-\alpha_{i-1}} r_i a_i^{\alpha_i}$; $1 \leq i \leq n$; is well defined. We have $|r'_i| = |r_i| + \alpha_i - \alpha_{i-1}$ and $\Sigma(r'_i) = \Sigma(r_i) + \alpha_i a_i - \alpha_{i-1}a_{i-1}$. By equation (7.7), for every $1 \leq i < n$, $|r'_i| = |r'_{i+1}|$. Moreover $r'_1 \ldots r'_n$ is a factor of w', and is non empty since $|r'_1 \ldots r'_n| \geq |r_1 \ldots r_n a_n| - 2$.

When we subtract (b) to (7.6), we get $\beta\Sigma(r'_i) = \beta\Sigma(r'_{i+1})$. Thus $\Sigma(r'_i) = \Sigma(r'_{i+1})$ for every $1 \leq i < n$, and w' has an additive-n-th-power $r'_1 \ldots r'_n$. □

$$s_{014}: \begin{cases} 0 \to \{0041141040110110040 11\} \\ 1 \to \{0041141040110110141 44\} \\ 2 \to \{0041141040100440441 44\} \\ 3 \to \{0041141040441440441 44\} \end{cases}$$

$$s_{025}: \begin{cases} 0 \to \{02200520220250552\} \\ 1 \to \{02252520220250552\} \\ 2 \to \{02255055200550552\} \\ 3 \to \{02255055252550552\} \end{cases}$$

Table 7.3 – Additive-cube-free substitutions from $\{0, 1, 2, 3\}^*$.

Theorem 7.6 can be used to find additive-square-free, additive-cube-free and additive-4th-power-free substitutions. However, we have few hopes to find an additive-square-free substitution, while additive-4th-powers are equivalent to abelian-4th-powers on binary words.

7.3.2 Additive-cube-free substitutions

We have checked by computer that every substitution in Table 7.2 respects the conditions of Theorem 7.6. Since there is an infinite additive-cube-free word on the alphabet $\{0, 1, 3, 4\}$, one can construct infinite additive-cube-free words on the alphabets $\{0, 1, 5\}$, $\{0, 1, 6\}$, $\{0, 1, 7\}$, $\{0, 2, 7\}$, $\{0, 3, 7\}$, $\{0, 1, 8\}$, $\{0, 3, 8\}$, $\{0, 1, 9\}$, $\{0, 2, 9\}$ and $\{0, 4, 9\}$. In our substitutions, each letter has at least two images. This clearly show that number of additive-cube-free words on these alphabets grows exponentially. For the alphabet $\{0, 1, 8\}$, we got 3 images of size 19 for each letter, giving the lower bound of $3^{1/19} = 1.059526\ldots$ for the growth rate. This bound is also a new lower bound for the growth rate of abelian-cube-free words on ternary alphabet. (The previous known bound was $2^{1/24} = 1.029302\ldots$ in [34].)

We conjecture that for every alphabet $A = \{0, i, j\}$ such that i and j are co-prime and $j \geq 6$, there exist an infinite additive-cube-free word on the alphabet A. The cases $\{0, 1, 2\}$, $\{0, 1, 3\}$, $\{0, 1, 4\}$ and $\{0, 2, 5\}$ are left open. Furthermore, it seems difficult to construct a very long word on the alphabet $\{0, 1, 2, 3\}$ avoiding additive cubes (the longest we got has size $\sim 1.4 \times 10^5$).

Question 7.7. Is there infinite additive-cube-free words on the following alphabets : $\{0, 1, 2, 3\}$, $\{0, 1, 4\}$ and $\{0, 2, 5\}$?

The substitutions in Table 7.3 also respect the conditions of Theorem 7.6, thus the existence of an infinite additive-cube-free word on $\{0, 1, 2, 3\}$ will imply the existence of infinite additive-cube-free words on $\{0, 1, 4\}$ and $\{0, 2, 5\}$.

Chapter 8

Avoiding k-binomial powers

In this chapter, we are dealing with another generalization of square-freeness and cube-freeness. We consider the k-binomial equivalence, which is an other refinement of the abelian equivalence. We prove that 2-binomial squares (resp. cubes) are avoidable over a 3-letter (resp. 2-letter) alphabet. The sizes of the alphabets are optimal. This chapter is based on the paper [31] (joint work with Michel Rigo and Pavel Salimov).

8.1 Introduction

The k-binomial equivalence relation is defined thanks to the binomial coefficient $\binom{u}{v}$ of two words u and v which is the number of times v occurs as a subsequence of u (meaning as a "scattered" subword). For more on these binomial coefficients, see for instance [111, Chap. 6]. Based on this classical notion, the m-binomial equivalence of two words has been recently introduced [133].

Definition 8.1. Let $m \in \mathbb{N} \cup \{+\infty\}$ and u, v be two words over the alphabet A. We let $A^{\leq m}$ denote the set of words of length at most m over A. We say that u and v are *m-binomially equivalent* if

$$\binom{u}{x} = \binom{v}{x}, \ \forall x \in A^{\leq m}.$$

We simply write $u \sim_m v$ if u and v are m-binomially equivalent. The word u is obtained as a permutation of the letters in v if and only if $u \sim_1 v$. In that case, we say that u and v are *abelian equivalent* and we write instead $u \sim_{\mathsf{ab}} v$. Note that if $u \sim_{k+1} v$, then $u \sim_k v$, for all $k \geq 1$.

Example. The four words 0101110, 0110101, 1001101 and 1010011 are 2-binomially equivalent. Let u be any of these four words. We have

$$\binom{u}{0} = 3, \ \binom{u}{1} = 4, \ \binom{u}{00} = 3, \ \binom{u}{01} = 7, \ \binom{u}{10} = 5, \ \binom{u}{11} = 6.$$

For instance, the word 0001111 is abelian equivalent to 0101110 but these two words are not 2-binomially equivalent. Let a be a letter. It is clear that $\binom{u}{aa}$ and $\binom{u}{a}$ carry the same information, *i.e.*, $\binom{u}{aa} = \binom{|u|_a}{2}$ where $|u|_a$ is the number of occurrences of a in u.

A *2-binomial square* (resp. *2-binomial cube*) is a non-empty word of the form xy where $x \sim_2 y$ (resp. $x \sim_2 y \sim_2 z$). Squares are avoidable over a 3-letter alphabet and abelian squares are avoidable over a 4-letter alphabet. Since 2-binomial equivalence lies between abelian equivalence and equality, the question is to determine whether or not 2-binomial squares are avoidable over a 3-letter alphabet. We answer positively to this question in Section 8.2. The fixed point of the morphism $g: 0 \mapsto 012, 1 \mapsto 02, 2 \mapsto 1$ avoids 2-binomial squares.

In a similar way, cubes are avoidable over a 2-letter alphabet and abelian squares are avoidable over a 3-letter alphabet. The question is to determine whether or not 2-binomial cubes are avoidable over a 2-letter alphabet. We also answer positively to this question in Section 8.3. The fixed point of the morphism $h: 0 \mapsto 001, 1 \mapsto 011$ avoids 2-binomial cubes.

The number of occurrences of a letter a in a word u will be denoted either by $\binom{u}{a}$ or $|u|_a$. Let $A = \{0, 1, \ldots, k\}$ be an alphabet. The *Parikh map* is an application $\Psi : A^* \to \mathbb{N}^{k+1}$ such that $\Psi(u) = (|u|_0, \dots, |u|_k)^T$. Note that we will deal with column vectors (when multiplying a square matrix with a column vector on its right). In particular, two words are abelian equivalent if and only if they have the same Parikh vector. The mirror of the word $u = u_1u_2 \cdots u_k$ is denoted by $\widetilde{u} = u_k \cdots u_2u_1$.

8.2 Avoiding 2-binomial squares over 3 letters

Let $A = \{0, 1, 2\}$ be a 3-letter alphabet. Let $g : A^* \to A^*$ be the morphism defined by

$$g : \begin{cases} 0 & \mapsto & 012 \\ 1 & \mapsto & 02 \\ 2 & \mapsto & 1 \end{cases} \quad \text{and thus, } g^2 : \begin{cases} 0 & \mapsto & 012021 \\ 1 & \mapsto & 0121 \\ 2 & \mapsto & 02. \end{cases}$$

It is prolongable on 0: $g(0)$ has 0 as a prefix. Hence the limit $\mathbf{x} = \lim_{n\to+\infty} g^n(0)$ is a well-defined infinite word

$$\mathbf{x} = g^\omega(0) = 012021012102012021020121 \cdots$$

which is a fixed point of g. Since the original work of Thue, this word $\mathbf{x}$ is well-known to avoid (usual) squares. It is sometimes referred to as the *ternary Thue–Morse word*. We will make use of the fact that $X = \{012, 02, 1\}$ is a prefix-code and thus an ω-code: Any finite word in X^* (resp. infinite word in X^ω) has a unique factorization as a product of elements in X. Let us make an obvious but useful observation.

Observation 8.2. *The factorization of* $\mathbf{x}$ *in terms of the elements in* X *permits to write* $\mathbf{x}$ *as*

$$\mathbf{x} = 0\,\alpha_1\,2\,\alpha_2\,0\,\alpha_3\,2\,\alpha_4\,0\,\alpha_5\,2\,\alpha_6\,0 \cdots$$

where, for all $i \geq 1$, $\alpha_i \in \{\varepsilon, 1\}$. *That is, the image of* $\mathbf{x}$ *by the morphism* $e : 0 \mapsto 0, 1 \mapsto \varepsilon, 2 \mapsto 2$ *(which erases all the* 1*'s) is* $e(\mathbf{x}) = (02)^\omega$.

The next property is well known. For example, it comes from the fact that the image of the ternary Thue–Morse word by the morphism $0 \mapsto 011, 1 \mapsto 01, 2 \mapsto 0$ is the Thue–Morse word. However, for the sake of completeness, we give a direct proof here.

Lemma 8.3. *A word* u *is a factor occurring in* $\mathbf{x}$ *if and only if* $\widetilde{u}$ *is a factor occurring in* $\mathbf{x}$.

Proof. We define the morphism $\widetilde{g} : A^* \to A^*$ by considering the mirror images of the images of the letters by g,

$$\widetilde{g} : \begin{cases} 0 & \mapsto & 210 \\ 1 & \mapsto & 20 \\ 2 & \mapsto & 1 \end{cases} \quad \text{and thus, } \widetilde{g}^2 : \begin{cases} 0 & \mapsto & 120210 \\ 1 & \mapsto & 1210 \\ 2 & \mapsto & 20. \end{cases}$$

Note that $\widetilde{g}$ is not prolongable on any letter. But the morphism $\widetilde{g}^2$ is prolongable on the letter 1. We consider the infinite word

$$\mathbf{y} = (\widetilde{g}^2)^\omega(1) = 121020121012021020120210121 0 \cdots .$$

If $v \in A^*$ is a non-empty word ending with $a \in A$, *i.e.*, $v = ua$ for some word $u \in A^*$, we denote by va^{-1} the word obtained by removing the suffix a from v. So $va^{-1} = u$.

For every words r and s we have $r = g^2(s) \Leftrightarrow \widetilde{r} = \widetilde{g}^2(\widetilde{s})$. Obviously, u is a factor occurring in $\mathbf{x}$ if and only if $\widetilde{u}$ is a factor occurring in $\mathbf{y}$.

On the other hand, $\widetilde{g}^2$ is a cyclic shift of g^2, since $g^2(a) = 0\widetilde{g}^2(a)0^{-1}$ for every $a \in \{0,1,2\}$. Thus u is a factor occurring in $\mathbf{x}$ if and only if u is a factor occurring in $\mathbf{y}$. To summarize, u is a factor occurring in $\mathbf{x}$ if and only if u is a factor occurring in $\mathbf{y}$, and u is a factor occurring in $\mathbf{y}$ if and only if $\widetilde{u}$ is a factor occurring in $\mathbf{x}$. This concludes the proof. ☐

We will be dealing with 2-binomial squares so, in particular, with abelian squares. The next lemma permit to "desubstitute", meaning that we are looking for the inverse image of a factor under the considered morphism.

Lemma 8.4. *Let $u, v \in A^*$ be two abelian equivalent non-empty words such that uv is a factor occurring in $\mathbf{x}$. There exists $u', v' \in A^*$ such that $u'v'$ is a factor of $\mathbf{x}$, and either:*

1. *$u = g(u')$ and $v = g(v')$;*
2. *or, $\widetilde{u} = g(v')$ and $\widetilde{v} = g(u')$.*

Proof. We will make an extensive use of Observation 8.2. Note that u and v must contain at least one 0 or one 2. Obviously $e(uv)$ is an abelian square of $(02)^\omega$, thus either $e(u) = e(v) = (02)^i$ or $e(u) = e(v) = (20)^i$ for an $i > 0$.

If $e(u) = e(v) = (02)^i$, then we have $u = a\,0 \cdots 2\,b$ and $v = c\,0 \cdots 2\,d$ with $a, bc, d \in \{\varepsilon, 1\}$. In this case, we deduce that u and v belongs to X^*. Otherwise stated, since uv is a factor of $\mathbf{x}$, there exists a factor $u'v'$ in $\mathbf{x}$ such that $g(u') = u$ and $g(v') = v$.

Otherwise we have $e(u) = e(v) = (20)^i$. Thanks to Lemma 8.3, $\widetilde{v}\widetilde{u}$ is a factor occurring in $\mathbf{x}$, and $e(\widetilde{u}) = e(\widetilde{v}) = (02)^i$. Thus we are reduced to the previous case, and there is a factor u', v' in $\mathbf{x}$ such that $g(u') = \widetilde{v}$ and $g(v') = \widetilde{u}$. ☐

Let u be a word. We set

$$\lambda_u := \binom{u}{01} - \binom{u}{12}.$$

When we use the desubstitution provided by the previous lemma, the shorter factors u' and v' derived from u and v keep properties from their ancestors.

Lemma 8.5. *Let $u, v \in A^*$ be two abelian equivalent non-empty words such that uv is a factor occurring in $\mathbf{x}$. Let u', v' be given by Lemma 8.4. If $\lambda_u = \lambda_v$, then u' and v' are abelian equivalent and $\lambda_{u'} = \lambda_{v'}$.*

Proof. If we are in the second situation described by Lemma 8.4, then $\widetilde{v}\widetilde{u}$ is also a factor occurring in $\mathbf{x}$. Obviously $\widetilde{v}$ and $\widetilde{u}$ are also abelian equivalent, $\lambda_{\widetilde{v}} = \lambda_{\widetilde{u}}$ and the case is reduced to the first situation.

Assume now w.l.o.g. that we are in the first situation, that is $u = g(u')$ and $v = g(v')$. First observe that we have, for all $a, b \in A$, $a \neq b$,

$$\binom{u'}{ab} = \binom{|u'|_a + |u'|_b}{2} - \binom{|u'|_a}{2} - \binom{|u'|_b}{2} - \binom{u'}{ba}. \tag{8.1}$$

Since $u = g(u')$, we derive that

$$\binom{u}{01} = |u'|_0 + \binom{u'}{00} + \binom{u'}{02} + \binom{u'}{12} + \binom{|u'|_0 + |u'|_1}{2} - \binom{|u'|_0}{2} - \binom{|u'|_1}{2} - \binom{u'}{01},$$

$$\begin{aligned}\binom{u}{12} &= |u'|_0 + \binom{u'}{00} + \binom{u'}{01} + \binom{|u'|_1 + |u'|_2}{2} - \binom{|u'|_1}{2} - \binom{|u'|_2}{2} - \binom{u'}{12} \\ &\quad + \binom{|u'|_0 + |u'|_2}{2} - \binom{|u'|_0}{2} - \binom{|u'|_2}{2} - \binom{u'}{02}.\end{aligned}$$

Hence

$$\begin{aligned}\lambda_u =& 2\left[\binom{u'}{02} - \binom{u'}{01} + \binom{u'}{12} - \binom{|u'|_2}{2}\right] \\ &+ \binom{|u'|_0 + |u'|_1}{2} - \binom{|u'|_1 + |u'|_2}{2} - \binom{|u'|_0 + |u'|_2}{2}.\end{aligned}$$

Similar relations holds for v.

Since u' and v' occur in $\mathbf{x}$, from Observation 8.2, we get

$$||u'|_0 - |u'|_2| \leq 1 \text{ and } ||v'|_0 - |v'|_2| \leq 1. \tag{8.2}$$

Since $u \sim_{\mathsf{ab}} v$, we have $|u|_1 = |v|_1$. Hence, from the definition of g, $|u'|_0 + |u'|_2 = |v'|_0 + |v'|_2$. In the same way, $|u|_2 = |v|_2$ implies that $|u'|_0 + |u'|_1 = |v'|_0 + |v'|_1$ or equivalently, $|u'|_1 - |v'|_1 = |v'|_0 - |u'|_0$. From the above relation and (8.2), we get

$$||v'|_0 - |u'|_0 + |u'|_2 - |v'|_2| \leq 2 \text{ and } |u'|_2 - |v'|_2 = |v'|_0 - |u'|_0.$$

Hence the difference of the following two Parikh vectors can only take three values

$$\Psi(u') - \Psi(v') \in \left\{ \begin{pmatrix}0\\0\\0\end{pmatrix}, \begin{pmatrix}1\\-1\\-1\end{pmatrix}, \begin{pmatrix}-1\\1\\1\end{pmatrix} \right\}.$$

To prove that u' and v' are abelian equivalent, we will rule out the last two possibilities.

By assumption, $\lambda_u = \lambda_v$. So this relation also holds modulo 2. Hence

$$\begin{aligned}&\binom{|u'|_0 + |u'|_1}{2} - \binom{|u'|_1 + |u'|_2}{2} - \binom{|u'|_0 + |u'|_2}{2} \\ \equiv\ &\binom{|v'|_0 + |v'|_1}{2} - \binom{|v'|_1 + |v'|_2}{2} - \binom{|v'|_0 + |v'|_2}{2} \pmod 2.\end{aligned}$$

Assume that we have

$$\Psi(u') - \Psi(v') = \begin{pmatrix}1\\-1\\-1\end{pmatrix}, i.e., \begin{array}{rcl} |u'|_0 + |u'|_1 &=& |v'|_0 + |v'|_1, \\ |u'|_0 + |u'|_2 &=& |v'|_0 + |v'|_2, \\ |u'|_1 + |u'|_2 &=& |v'|_1 + |v'|_2 - 2. \end{array}$$

This leads to a contradiction because then

$$\binom{|u'|_1 + |u'|_2}{2} \not\equiv \binom{|v'|_1 + |v'|_2}{2} \pmod 2.$$

Indeed, it is easily seen that $\binom{4n}{2} \equiv 0 \pmod 2$, $\binom{4n+1}{2} \equiv 0 \pmod 2$, $\binom{4n+2}{2} \equiv 1 \pmod 2$ and $\binom{4n+3}{2} \equiv 1 \pmod 2$.

The case $\Psi(u') - \Psi(v') = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$ is handled similarly. So we can assume now that $\Psi(u') = \Psi(v')$, that is $u' \sim_{\mathsf{ab}} v'$. It remains to prove that $\lambda_{u'} = \lambda_{v'}$. By assumption $\lambda_u = \lambda_v$, and from the above formula describing λ_u (resp. λ_v) we get

$$\binom{u'}{02} - \binom{u'}{01} + \binom{u'}{12} = \binom{v'}{02} - \binom{v'}{01} + \binom{v'}{12}.$$

To conclude that $\lambda_{u'} = \lambda_{v'}$, we should simply show that $\binom{u'}{02} = \binom{v'}{02}$. But $u'v'$ is a factor occurring in $\mathbf{x}$ (from Observation 8.2, when discarding the 1's with just alternate 0's and 2's) and $u' \sim_{\mathsf{ab}} v'$. This concludes the proof. □

Theorem 8.6. *The word* $\mathbf{x} = g^\omega(0) = 0120210121020120210201 21\cdots$ *avoids* 2*-binomial squares.*

Proof. Assume to the contrary that $\mathbf{x}$ contains a 2-binomial square uv where u and v are 2-binomially equivalent. In particular, u and v are abelian equivalent and moreover $\lambda_u = \lambda_v$. We can therefore apply iteratively Lemma 8.4 and the above lemma to words of decreasing lengths and get finally a repetition aa with $a \in A$ in $\mathbf{x}$. But $\mathbf{x}$ does not contain any such factor. □

Remark 8.7. The fixed point of g is 2-binomial-square free, but g is not 2-binomial-square-free, that is the image of a 2-binomial-square-free word may contain a 2-binomial-square (*e.g.*, $g(010) = 01202012$ contains the square 2020).

8.3 Avoiding 2-binomial cubes over 2 letters

Consider the morphism $h : 0 \mapsto 001$ and $h : 1 \mapsto 011$. In this section, we show that h is 2-binomial-cube-free, that is for every 2-binomial-cube free binary word w, $h(w)$ is 2-binomial-cube-free. As a direct corollary, we get that the fixed point of h,

$$\mathbf{z} = h^\omega(0) = 001001011001001011001011011\cdots$$

avoids 2-binomial cubes.

Let u be a word over $\{0,1\}$. The *extended Parikh vector* of u is

$$\Psi_2(u) = \left(|u|_0, |u|_1, \binom{u}{00}, \binom{u}{01}, \binom{u}{10}, \binom{u}{11}\right)^T.$$

Observe that two words u and v are 2-binomially equivalent if and only if $\Psi_2(u) = \Psi_2(v)$.

Consider the matrix M_h given by

$$M_h = \begin{pmatrix} 2 & 1 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 & 0 \\ 1 & 0 & 4 & 2 & 2 & 1 \\ 2 & 2 & 2 & 4 & 1 & 2 \\ 0 & 0 & 2 & 1 & 4 & 2 \\ 0 & 1 & 1 & 2 & 2 & 4 \end{pmatrix}.$$

One can check that M_h is invertible. We will make use of the following observations:

Proposition 8.8. *For every $u \in \{0,1\}^*$,*

$$\Psi_2(h(u)) = M_h \Psi_2(u).$$

Proposition 8.9. *Let $u = 1x$ and $u' = x1$ be two words over $\{0,1\}$. We have $|u|_0 = |u'|_0$, $|u|_1 = |u'|_1$,*

$$\binom{u}{00} = \binom{u'}{00},\ \binom{u}{11} = \binom{u'}{11},\ \binom{u'}{01} = \binom{u}{01} + |u|_0,\ \binom{u'}{10} = \binom{u}{10} - |u|_0.$$

In particular, if $1x \sim_2 1y$, then $x1 \sim_2 y1$. Similar relations hold for $0x$ and $x0$. In particular, if $x0 \sim_2 y0$, then $0x \sim_2 0y$.

Let $x, y \in \{0,1\}$. We set $\delta_{x,y} = 1$, if $x = y$; and $\delta_{x,y} = 0$, otherwise.

Lemma 8.10. *Let p', q' and r' be binary words, and let $a, b \in \{0,1\}$. Let $p = h(p')\,0$, $q = a\,1\,h(q')\,0\,b$ and $r = 1\,h(r')$. Then either $p \not\sim_2 q$ or $p \not\sim_2 r$.*

Proof. Assume, for the sake of contradiction, that $p \sim_2 q \sim_2 r$. Then $|p'| = |q'| + 1 = |r'| = n$. The following relations can mostly be derived from the coefficients of M_h (we also have to take into account the extra suffix 0 of p, respectively the extra prefix 1 in r):

$$\binom{p}{01} = 2\binom{p'}{0} + 2\binom{p'}{1} + 2\binom{p'}{00} + 4\binom{p'}{01} + \binom{p'}{10} + 2\binom{p'}{11},$$

$$\binom{p}{10} = \binom{p'}{0} + 2\binom{p'}{1} + 2\binom{p'}{00} + \binom{p'}{01} + 4\binom{p'}{10} + 2\binom{p'}{11},$$

$$\Rightarrow \binom{p}{01} - \binom{p}{10} = \binom{p'}{0} + 3\binom{p'}{01} - 3\binom{p'}{10};$$

$$\binom{r}{01} = 2\binom{r'}{0} + 2\binom{r'}{1} + 2\binom{r'}{00} + 4\binom{r'}{01} + \binom{r'}{10} + 2\binom{r'}{11},$$

$$\binom{r}{10} = 2\binom{r'}{0} + \binom{r'}{1} + 2\binom{r'}{00} + \binom{r'}{01} + 4\binom{r'}{10} + 2\binom{r'}{11},$$

$$\Rightarrow \binom{r}{01} - \binom{r}{10} = \binom{r'}{1} + 3\binom{r'}{01} - 3\binom{r'}{10}.$$

We also get the following relations:

$$\begin{aligned}
\binom{q}{01} &= 2\binom{q'}{0} + 2\binom{q'}{1} + 2\binom{q'}{00} + 4\binom{q'}{01} + \binom{q'}{10} + 2\binom{q'}{11} \\
&\quad + \delta_{a,0}\left[1 + \binom{q'}{0} + 2\binom{q'}{1} + \delta_{b,1}\right] + \delta_{b,1}\left[1 + 2\binom{q'}{0} + \binom{q'}{1}\right],
\end{aligned}$$

$$\begin{aligned}
\binom{q}{10} &= 3\binom{q'}{0} + 3\binom{q'}{1} + 2\binom{q'}{00} + \binom{q'}{01} + 4\binom{q'}{10} + 2\binom{q'}{11} + 1 \\
&\quad + \delta_{a,1}\left[1 + \delta_{b,0} + 2\binom{q'}{0} + \binom{q'}{1}\right] + \delta_{b,0}\left[1 + \binom{q'}{0} + 2\binom{q'}{1}\right] \\
&= (6 - 2\delta_{a,0} - \delta_{b,1})\binom{q'}{0} + (6 - \delta_{a,0} - 2\delta_{b,1})\binom{q'}{1} + 4 - 2\delta_{a,0} - 2\delta_{b,1} \\
&\quad + \delta_{a,0}\delta_{b,1} + 2\binom{q'}{00} + \binom{q'}{01} + 4\binom{q'}{10} + 2\binom{q'}{11}.
\end{aligned}$$

Where for the last equality, we have used the fact that $\delta_{a,1} = 1 - \delta_{a,0}$ and $\delta_{b,0} = 1 - \delta_{b,1}$. Finally, we obtain

$$\binom{q}{01} - \binom{q}{10} = (-4+3\delta_{a,0}+3\delta_{b,1})\left[\binom{q'}{0}+\binom{q'}{1}\right]+3\binom{q'}{01}-3\binom{q'}{10}-4+3\delta_{a,0}+3\delta_{b,1}.$$

Since $p \sim_2 q \sim_2 r$, we have $\binom{p}{10} - \binom{p}{01} = \binom{q}{10} - \binom{q}{01} = \binom{r}{10} - \binom{r}{01}$. In particular, these equalities modulo 3 give

$$\binom{p'}{0} \equiv \binom{r'}{1} \equiv 2\left[\binom{q'}{0} + \binom{q'}{1} + 1\right] \equiv 2n \pmod 3. \tag{8.3}$$

Now, we take into account the fact that p and r are abelian equivalent to get a contradiction. Since $p = h(p')\,0$ and $r = 1\,h(r')$, we get

$$\begin{pmatrix}|p|_0\\|p|_1\end{pmatrix} = \begin{pmatrix}2 & 1\\1 & 2\end{pmatrix}\begin{pmatrix}|p'|_0\\|p'|_1\end{pmatrix} + \begin{pmatrix}1\\0\end{pmatrix}, \quad \begin{pmatrix}|r|_0\\|r|_1\end{pmatrix} = \begin{pmatrix}2 & 1\\1 & 2\end{pmatrix}\begin{pmatrix}|r'|_0\\|r'|_1\end{pmatrix} + \begin{pmatrix}0\\1\end{pmatrix}.$$

Hence, we obtain

$$\begin{pmatrix}|p|_0 - |r|_0\\|p|_1 - |r|_1\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix} = \begin{pmatrix}2 & 1\\1 & 2\end{pmatrix}\begin{pmatrix}|p'|_0 - |r'|_0\\|p'|_1 - |r'|_1\end{pmatrix} + \begin{pmatrix}1\\-1\end{pmatrix}.$$

We derive that $|p'|_0 - |r'|_0 = -1$ and $|p'|_1 - |r'|_1 = 1$. Recalling that $|p'|_0 + |p'|_1 = n$. If we subtract the last two equalities, we get $|p'|_0 + |r'|_1 = n - 1$. From (8.3), we know that $|p'|_0 \equiv |r'|_1 \pmod 3$. Hence $2|p'|_0 \equiv n - 1 \pmod 3$ and thus

$$|p'|_0 \equiv 2n - 2 \pmod 3.$$

This contradicts the fact again given by (8.3) that $|p'|_0 \equiv 2n \pmod 3$. □

Similarly, one get the following lemma.

Lemma 8.11. *Let p', q' and r' be binary words, and let $a, b \in \{0,1\}$. Let $p = h(p')\,0\,a$, $q = 1\,h(q')\,0$ and $r = b\,1\,h(r')$. Then either $p \not\sim_2 q$ or $p \not\sim_2 r$.*

Proof. Assume, for the sake of contradiction, that $p \sim_2 q \sim_2 r$. Then $|p'| = |q'| = |r'| = n$. Taking into account the special form of p and q, we get

$$\binom{p}{01} = 2\binom{p'}{0}+2\binom{p'}{1}+2\binom{p'}{00}+4\binom{p'}{01}+\binom{p'}{10}+2\binom{p'}{11}+\delta_{a,1}\left(1+2\binom{p'}{0}+\binom{p'}{1}\right),$$

$$\binom{p}{10} = \binom{p'}{0} + 2\binom{p'}{1} + 2\binom{p'}{00} + \binom{p'}{01} + 4\binom{p'}{10} + 2\binom{p'}{11} + \delta_{a,0}\left(\binom{p'}{0} + 2\binom{p'}{1}\right),$$

$$\binom{q}{01} = 2\binom{q'}{0} + 2\binom{q'}{1} + 2\binom{q'}{00} + 4\binom{q'}{01} + \binom{q'}{10} + 2\binom{q'}{11},$$

$$\binom{q}{10} = 3\binom{q'}{0} + 3\binom{q'}{1} + 2\binom{q'}{00} + \binom{q'}{01} + 4\binom{q'}{10} + 2\binom{q'}{11} + 1.$$

Hence, we get

$$\binom{p}{01} - \binom{p}{10} = -2\binom{p'}{1} + 3\binom{p'}{01} - 3\binom{p'}{10} + \delta_{a,1}\left(1 + 3\binom{p'}{0} + 3\binom{p'}{1}\right),$$

$$\binom{q}{01}-\binom{q}{10}=-\binom{q'}{0}-\binom{q'}{1}+3\binom{q'}{01}-3\binom{q'}{10}-1.$$

Since, $p \sim_2 q$, the last two relations evaluated modulo 3 give

$$|p'|_1+\delta_{a,1}\equiv 2n+2 \pmod 3. \tag{8.4}$$

Similarly, the form of r gives the following relations

$$\binom{r}{01}=2\binom{r'}{0}+2\binom{r'}{1}+2\binom{r'}{00}+4\binom{r'}{01}+\binom{r'}{10}+2\binom{r'}{11}+\delta_{b,0}\left(1+\binom{r'}{0}+2\binom{r'}{1}\right),$$

$$\binom{r}{10}=2\binom{r'}{0}+\binom{r'}{1}+2\binom{r'}{00}+\binom{r'}{01}+4\binom{r'}{10}+2\binom{r'}{11}+\delta_{b,1}\left(2\binom{r'}{0}+\binom{r'}{1}\right),$$

$$\binom{r}{01}-\binom{r}{10}=-2\binom{r'}{0}+3\binom{r'}{01}-3\binom{r'}{10}+\delta_{b,0}\left(1+3\binom{r'}{0}+3\binom{r'}{1}\right)$$

Since, $p \sim_2 r$, the last two relations evaluated modulo 3 give

$$|p'|_1+\delta_{a,1}\equiv |r'|_0+\delta_{b,0} \pmod 3. \tag{8.5}$$

Now, we take into account the fact that p, q and r are abelian equivalent to get a contradiction. The following two vectors are equal:

$$\begin{pmatrix}|p|_0\\|p|_1\end{pmatrix}=\begin{pmatrix}2&1\\1&2\end{pmatrix}\begin{pmatrix}|p'|_0\\|p'|_1\end{pmatrix}+\begin{pmatrix}1+\delta_{a,0}\\\delta_{a,1}\end{pmatrix},\ \begin{pmatrix}|r|_0\\|r|_1\end{pmatrix}=\begin{pmatrix}2&1\\1&2\end{pmatrix}\begin{pmatrix}|r'|_0\\|r'|_1\end{pmatrix}+\begin{pmatrix}\delta_{b,0}\\1+\delta_{b,1}\end{pmatrix}.$$

We derive easily that

$$|p'|_1-|r'|_1=\ 1+\delta_{a,0}-\delta_{b,0}.$$

On the one hand, using the latter relation and (8.5)

$$|r'|_1+1+\delta_{a,0}-\delta_{b,0}+\delta_{a,1}=|p'|_1+\delta_{a,1}\equiv|r'|_0+\delta_{b,0} \pmod 3$$

Replacing $|r'|_0$ by $n-|r'|_1$, we get $2|r'|_1+2\equiv n+2\delta_{b,0} \pmod 3$, or equivalently

$$|r'|_1+1\equiv 2n+\delta_{b,0} \pmod 3.$$

On the other hand, using (8.4),

$$|r'|_1+1+\delta_{a,0}-\delta_{b,0}+\delta_{a,1}=|p'|_1+\delta_{a,1}\equiv 2n+2 \pmod 3$$

and thus,

$$|r'|_1\equiv 2n+\delta_{b,0} \pmod 3.$$

We get a contradiction, $2n+\delta_{b,0}$ should congruent to both $|r'|_1$ and $|r'|_1+1$ modulo 3. □

We are ready to prove the main theorem of this section.

Theorem 8.12. *Let $h: 0\mapsto 001, 1\mapsto 011$. For every 2-binomial-cube-free word $w\in\{0,1\}^*$, $h(w)$ is 2-binomial-cube-free.*

Proof. Let w be a 2-binomial-cube-free binary word. Assume that $h(w) = z_0 \dots z_{3|w|-1}$ contains a 2-binomial cube pqr occurring in position i, *i.e.*, $p \sim_2 q \sim_2 r$ and $w = w'\, p\, q\, r\, w''$, where $|w'| = i$. We consider three cases depending on the size of p modulo 3.

As a first case, assume that $|p| = 3n$. We consider three sub-cases depending on the position i modulo 3.

1.a) Assume that $i \equiv 2 \pmod 3$. Then p, q, r have 1 as a prefix and the letter following r in $h(w)$ is the symbol $z_{i+9n} = 1$. Hence, the word $1^{-1}pqr1$ occurs in $h(w)$ in position $i + 1$ and it is again a 2-binomial cube. Indeed, thanks to Proposition 8.9, we have $1^{-1}p1 \sim_2 1^{-1}q1 \sim_2 1^{-1}r1$. This case is thus reduced to the case where $i \equiv 0 \pmod 3$.

1.b) Assume that $i \equiv 1 \pmod 3$. Then p, q, r have 0 as a suffix and the letter preceding p in $h(w)$ is the symbol $z_{i-1} = 0$. Hence, the word $0pqr0^{-1}$ occurs in $h(w)$ in position $i - 1$ and it is also a 2-binomial cube. Thanks to Proposition 8.9, we have $0p0^{-1} \sim_2 0q0^{-1} \sim_2 0r0^{-1}$. Again this case is reduced to the case where $i \equiv 0 \pmod 3$.

1.c) Assume that $i \equiv 0 \pmod 3$. In this case, we can desubstitute: there exist three words p', q', r' of length n such that $h(p') = p$, $h(q') = q$, $h(r') = r$ and $p'q'r'$ is a factor occurring in w. We have $\Psi_2(p) = \Psi_2(q) = \Psi_2(r)$. By Proposition 8.8, and since M_h is invertible, we have $\Psi_2(p') = \Psi_2(q') = \Psi_2(r')$, meaning that w contains a 2-binomial cube $p'q'r'$.

As a second case, assume that $|p| = 3n + 1$. In this case, one of p, q and r occur in position 0 modulo 3, one in position 1 modulo 3, and one in position 2 modulo 3. Suppose w.l.o.g. that p occur in position 0 modulo 3, and q in position 1 modulo 3. Then there are three factors p', q' and r' in w, and $a, b \in \{0, 1\}$ such that $p = h(p')\,0$, $q = a\,1\,h(q')\,0\,b$ and $r = 1\,h(r')$. By Lemma 8.10, this is impossible.

For the final case, assume that $|p| = 3n + 2$. In this case again, one of p, q and r occur in position 0 modulo 3, one in position 1 modulo 3, and one in position 2 modulo 3. Suppose w.l.o.g. that p occur in position 0 modulo 3, and q in position 1 modulo 3. Then there are three factors p', q' and r' in w, and $a, b \in \{0, 1\}$ such that $p = h(p')\,0\,a$, $q = 1\,h(q')\,0$ and $r = b\,1\,h(r')$. By Lemma 8.11, this is impossible. $\square$

Corollary 8.13. *The infinite word* $\mathbf{z} = 001001011\cdots$ *fixed point of* $h : 0 \mapsto 001, 1 \mapsto 011$ *avoids* 2*-binomial cubes.*

Part III

Two two-dimensional problems

Chapter 9

The domination number of grids

This chapter presents a joint work with Daniel Gonçalves, Alexandre Pinlou and Stéphan Thomassé, published in *SIAM Journal on Discrete Mathematics* [10]. We conclude the calculation of the domination number of all $n \times m$ grid graphs. Indeed, we prove Chang's conjecture saying that for every $16 \le n \le m$, $\gamma(G_{n,m}) = \left\lfloor \frac{(n+2)(m+2)}{5} \right\rfloor - 4$.

9.1 Introduction

A *dominating set* in a graph G is a subset of vertices S such that every vertex in $V(G) \setminus S$ is a neighbour of some vertex of S. The *domination number* of G is the minimum size of a dominating set of G. We denote it by $\gamma(G)$. This chapter is devoted to the calculation of the domination number of complete grids.

The notation $[i]$ denotes the set $\{1, 2, \dots, i\}$. If w is a word on the alphabet A, $w[i]$ is the i-th letter of w, and for every a in A, $|w|_a$ denotes the number of occurrences of a in w (i.e. $|\{i \in \{1, \dots, |w|\} : w[i] = a\}|$). For a vertex v, $N[v]$ denotes the closed neighbourhood of v (i.e. the set of neighbours of v and v itself). For a subset of vertices S of a vertex set V of a graph, we denote by $N[S] = \bigcup_{v \in S} N[v]$. Note that D is a dominating set of G if and only if $N[D] = V(G)$. Let $G_{n,m}$ be the $n \times m$ complete grid, i.e. the vertex set of $G_{n,m}$ is $V_{n,m} := [n] \times [m]$, and two vertices (i, j) and (k, l) are adjacent if $|k - i| + |l - j| = 1$. The couple $(1, 1)$ denotes the bottom-left vertex of the grid and the couple (i, j) denotes the vertex of the i-th column and the j-th row. We will always assume in this chapter that $n \le m$. Let us illustrate our purpose by an example of a dominating set of the complete grid $G_{24,24}$ on Figure 9.1.

The first results on the domination number of grids were obtained about 30 years ago with the exact values of $\gamma(G_{2,n})$, $\gamma(G_{3,n})$, and $\gamma(G_{4,n})$ found by Jacobson and Kinch [91] in 1983. In 1993, Chang and Clark [52] found those of $\gamma(G_{5,n})$ and $\gamma(G_{6,n})$. These results were obtained analytically. Chang [53] devoted his PhD thesis to study the domination number of grids; he conjectured that this invariant behaves well provided that n is large enough. Specifically, Chang conjectured the following:

Conjecture 9.1 ([53])**.** For every $16 \le n \le m$,

$$\gamma(G_{n,m}) = \left\lfloor \frac{(n+2)(m+2)}{5} \right\rfloor - 4.$$

Observe that for instance, this formula would give 131 for the domination number of the grid in Figure 9.1. To motivate his bound, Chang proposed some constructions of dominating sets achieving the upper bound:

Lemma 9.2 ([53])**.** *For every* $8 \le n \le m$*,*

$$\gamma(G_{n,m}) \le \left\lfloor \frac{(n+2)(m+2)}{5} \right\rfloor - 4$$

Later, some algorithms based on dynamic programming were designed to compute a lower bound of $\gamma(G_{n,m})$. There were numerous intermediate results found for $\gamma(G_{n,m})$ for small values of n and m (see [54, 83, 141] for details). In 1995, Hare, Hedetniemi and Hare [83] gave a polynomial time algorithm to compute $\gamma(G_{n,m})$ when n is fixed. Nevertheless, this algorithm is not usable in practice when n hangs over 20. Fisher [76] developed the idea of searching for periodicity in the dynamic programming algorithms and using this technique, he found the exact values of $\gamma(G_{n,m})$ for all $n \leq 21$. We recall these values for the sake of completeness.

Theorem 9.3 ([76]). *For all $n \leq m$ and $n \leq 21$, we have:*

$$\gamma(G_{n,m}) = \begin{cases} \lceil \frac{m}{3} \rceil & \text{if } n = 1 \\ \lceil \frac{m+1}{2} \rceil & \text{if } n = 2 \\ \lceil \frac{3m+1}{4} \rceil & \text{if } n = 3 \\ m+1 & \text{if } n = 4 \text{ and } m = 5,6,9 \\ m & \text{if } n = 4 \text{ and } m \neq 5,6,9 \\ \lceil \frac{6m+4}{5} \rceil - 1 & \text{if } n = 5 \text{ and } m = 7 \\ \lceil \frac{6m+4}{5} \rceil & \text{if } n = 5 \text{ and } m \neq 7 \\ \lceil \frac{10m+6}{7} \rceil - 1 & \text{if } n = 6 \text{ and } m \equiv_7 1 \\ \lceil \frac{10m+6}{7} \rceil & \text{if } n = 6 \text{ and } m \not\equiv_7 1 \\ \lceil \frac{5m+1}{3} \rceil & \text{if } n = 7 \\ \lceil \frac{15m+7}{8} \rceil & \text{if } n = 8 \\ \lceil \frac{23m+10}{11} \rceil & \text{if } n = 9 \\ \lceil \frac{30m+15}{13} \rceil - 1 & \text{if } n = 10 \text{ and } m \equiv_{13} 10 \text{ or } m = 13,16 \\ \lceil \frac{30m+15}{13} \rceil & \text{if } n = 10 \text{ and } m \not\equiv_{13} 10 \text{ and } m \neq 13,16 \\ \lceil \frac{38m+22}{15} \rceil - 1 & \text{if } n = 11 \text{ and } m = 11,18,20,22,33 \\ \lceil \frac{38m+22}{15} \rceil & \text{if } n = 11 \text{ and } m \neq 11,18,20,22,33 \\ \lceil \frac{80m+38}{29} \rceil & \text{if } n = 12 \\ \lceil \frac{98m+54}{33} \rceil - 1 & \text{if } n = 13 \text{ and } m \equiv_{33} 13,16,18,19 \\ \lceil \frac{98m+54}{33} \rceil & \text{if } n = 13 \text{ and } m \not\equiv_{33} 13,16,18,19 \\ \lceil \frac{35m+20}{11} \rceil - 1 & \text{if } n = 14 \text{ and } m \equiv_{22} 7 \\ \lceil \frac{35m+20}{11} \rceil & \text{if } n = 14 \text{ and } m \not\equiv_{22} 7 \\ \lceil \frac{44m+28}{13} \rceil - 1 & \text{if } n = 15 \text{ and } m \equiv_{26} 5 \\ \lceil \frac{44m+28}{13} \rceil & \text{if } n = 15 \text{ and } m \not\equiv_{26} 5 \\ \lfloor \frac{(n+2)(m+2)}{5} \rfloor - 4 & \text{if } n \geq 16 \end{cases}$$

Note that these values are obtained by specific formulas for every $1 \leq n \leq 15$ and by the formula of Conjecture 9.1 for every $16 \leq n \leq 21$. This proves Chang's conjecture for all values $16 \leq n \leq 21$.

In 2004, Conjecture 9.1 has been confirmed up to an additive constant:

Theorem 9.4 (Guichard [82]). *For every $16 \leq n \leq m$,*

$$\gamma(G_{n,m}) \geq \left\lfloor \frac{(n+2)(m+2)}{5} \right\rfloor - 9.$$

In this chapter, we prove Chang's conjecture, hence finishing the computation of $\gamma(G_{n,m})$. We adapt Guichard's ideas to improve the additive constant from -9 to -4 when $24 \leq n \leq m$. Cases $n = 22$ and $n = 23$ can be proved in a couple of hours using Fisher's method (described in [76]) on a modern computer.

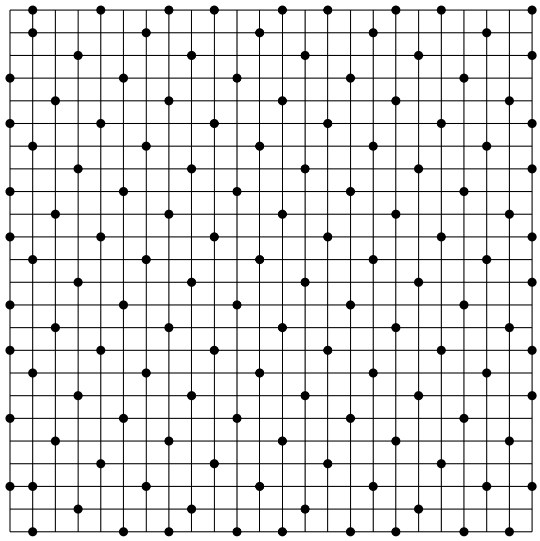


Figure 9.1 – Example of a set of size 131 dominating the grid $G_{24,24}$

They can be also proved by a slight improvement of the technique presented in the next section.

9.2 Values of $\gamma(G_{n,m})$ when $24 \leq n \leq m$

Our method follows the idea of Guichard [82]. A slight improvement is enough to give the exact bound.

A vertex of the grid $G_{n,m}$ dominates at most 5 vertices (its four neighbours and itself). It is then clear that $\gamma(G_{n,m}) \geq \frac{n \times m}{5}$. The previous inequality would become an equality if there would be a dominating set D such that every vertex of $G_{n,m}$ is dominated only once, and all vertices of D are of degree 4 (i.e. dominates exactly 5 vertices); in this case, we would have $5 \times |D| - n \times m = 0$. This is clearly impossible (e.g. to dominate the corners of the grid, we need vertices of degree at most 3). Therefore, our goal is to find a dominating set D of $G_{n,m}$ such that the difference $5 \times |D| - n \times m$ is the smallest.

Let S be a subset of $V(G_{n,m})$. The *loss* of S is $\ell(S) = 5 \times |S| - |N[S]|$.

Proposition 9.5. *The following properties of the loss function are straightforward:*

(i) For every S, $\ell(S) \geq 0$ (positivity),

(ii) If $S_1 \cap S_2 = \emptyset$, then $\ell(S_1 \cup S_2) = \ell(S_1) + \ell(S_2) + |N[S_1] \cap N[S_2]|$,

(iii) If $S' \subseteq S$, then $\ell(S') \leq \ell(S)$ (increasing function),

(iv) If $S_1 \cap S_2 = \emptyset$, then $\ell(S_1 \cup S_2) \geq \ell(S_1) + \ell(S_2)$ (super-additivity).

Let us denote by $\ell_{n,m}$ the minimum of $\ell(D)$ when D is a dominating set of $G_{n,m}$.

Lemma 9.6. $\gamma(G_{n,m}) = \left\lceil \frac{n \times m + \ell_{n,m}}{5} \right\rceil$

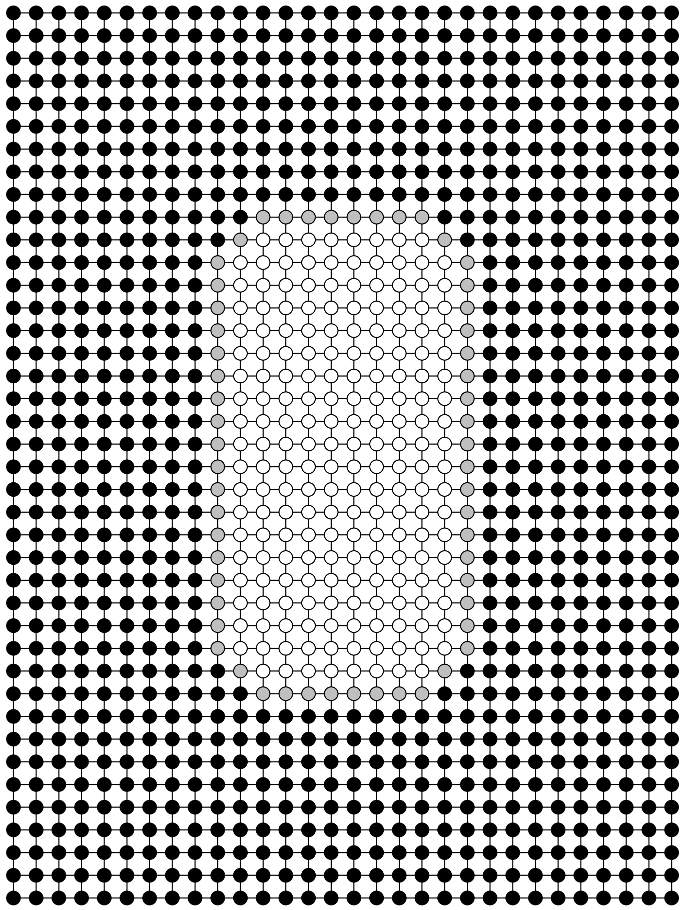


Figure 9.2 – The graph $G_{30,40}$. The set $I(B_{30,40})$ is the set of vertices filled in black. The set $B_{30,40}$ is the set of vertices filled in black or in gray.

Proof. If D is a dominating set of $G_{n,m}$, then $\ell(D) = 5 \times |D| - |N[D]| = 5 \times |D| - n \times m$. Hence, by minimality of $\ell_{n,m}$ and $\gamma(G_{n,m})$, we have $\ell_{n,m} = 5 \times \gamma(G_{n,m}) - n \times m$. □

Our aim is to get a lower bound for $\ell_{n,m}$. As the reader can observe in Figure 9.1, the loss is concentrated on the border of the grid. We now analyse more carefully the loss generated by the border of thickness 10.

We define the border $B_{n,m} \subseteq V_{n,m}$ of $G_{n,m}$ as the set of vertices (i,j) where $i \leq 10$, or $j \leq 10$, or $i > n-10$, or $j > m-10$ to which we add the four vertices $(11,11), (11,m-10), (n-10,11), (n-10,m-10)$. Given a subset $S \subseteq V$, let $I(S)$ be the *internal vertices* of S, i.e. $I(S) = \{v \in S : N[v] \subseteq S\}$. These sets are illustrated in Figure 9.2. We will compute $b_{n,m} = \min_D \ell(D)$, where D is a subset of $B_{n,m}$ and dominates $I(B_{n,m})$, i.e. $D \subseteq B_{n,m}$ and $I(B_{n,m}) \subseteq N[D]$. Observe that this lower bound $b_{n,m}$ is a lower bound of $\ell_{n,m}$. Indeed, for every dominating set D of $G_{n,m}$, the set $D' := D \cap B_{n,m}$ dominates $I(B_{n,m})$, hence $b_{n,m} \leq \ell(D') \leq \ell(D)$. In the remainder, we will compute $b_{n,m}$ and we will show that $b_{n,m} = \ell_{n,m}$.

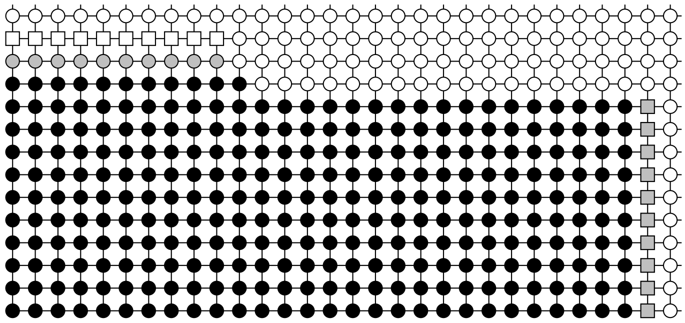


Figure 9.3 – The set P_{19} (black and gray), the set of input vertices (gray circles) and the set of output vertices (gray squares).

In the following, we split the border $B_{n,m}$ in four parts: O_{m-12}, P_{n-12}, Q_{m-12} and R_{n-12}, which are defined just below.

For $p \geq 12$, let $P_p \subset B_{n,m}$ defined as follows : $P_p = ([10] \times \{12\}) \cup ([11] \times \{11\}) \cup ([p] \times [10])$. We define the ***input vertices*** of P_p as $[10] \times \{12\}$ and the ***output vertices*** of P_p as $\{p\} \times [10]$. The set P_p, illustrated for $p = 19$ in Figure 9.3, corresponds to the set of black and gray vertices. The input vertices are the gray circles, and the output vertices are the gray squares. Recall that in our drawing conventions, the vertex $(1,1)$ is the bottom-left vertex and hence the vertex (i,j) is in the i^{th} column from the left and in the j^{th} row from the bottom.

For $n, m \in \mathbb{N}^*$, let $f_{n,m} : [n] \times [m] \to [m] \times [n]$ be the bijection such that $f_{n,m}(i,j) = (j, n-i+1)$. It is clear that the set $B_{n,m}$ is the disjoint union of the following four sets depicted in Figure 9.4: P_{n-12}, $Q_{m-12} = f_{n,m}(P_{m-12})$, $R_{n-12} = f_{m,n} \circ f_{n,m}(P_{n-12})$ and $O_{m-12} = f_{n,m}^{-1}(P_{m-12})$. Similarly to P_{n-12}, the sets O_{m-12}, Q_{n-12} and R_{m-12} have input and output vertices. For instance, the output vertices of Q_{m-12} correspond in Figure 9.3 to the white squares. Every set playing a symmetric role, we now focus on P_{n-12}.

Given a subset S of $V_{n,m}$, let the labelling $\phi_S : V_{n,m} \to \{0,1,2\}$ be such that

$$\phi_S(i,j) = \begin{cases} 0 & \text{if } (i,j) \in S \\ 1 & \text{if } (i,j) \in N[S] \setminus S \\ 2 & \text{otherwise} \end{cases}$$

Note that ϕ_S is such that any two adjacent vertices in $G_{n,m}$ cannot be labelled 0 and 2.

Given $p \geq 12$ and a set $S \subseteq P_p$, the *input word* (resp. *output word*) of S for P_p, denoted by $w^{in}(S)$ (resp. $w_p^{out}(S)$), is the ten letters word on the alphabet $\{0,1,2\}$ obtained by reading ϕ_S on the input vertices (resp. output vertices) of P_p. More precisely, its i^{th} letter is $\phi_S(i,12)$ (resp. $\phi_S(p,i)$). Similarly, O_p, Q_p and R_p have also input and output words. For example, the output word of $S \subseteq O_p$ for O_p is $w_p^{out}(f_{n,m}(S))$.

According to the definition of ϕ, the input and output words belong to the set $\mathcal{W}$ of ten letters words on $\{0,1,2\}$ which avoid 02 and 20. The number of

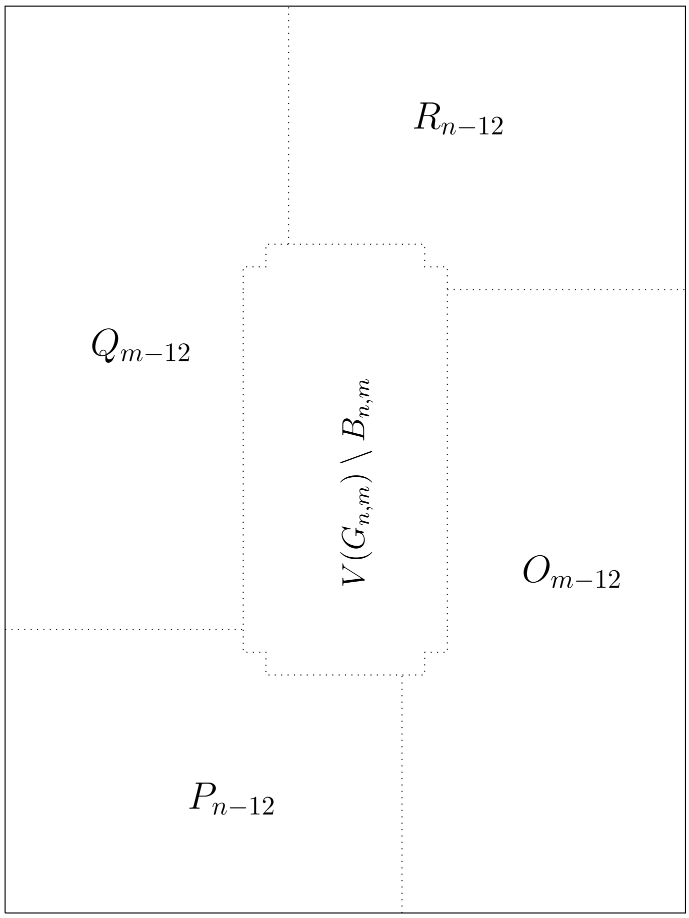


Figure 9.4 – The sets O_{m-12}, P_{n-12}, Q_{m-12} and R_{n-12}.

k-digits trinary numbers without 02 or 20 is given by the following formula [76]:

$$\frac{(1+\sqrt{2})^{k+1}+(1-\sqrt{2})^{k+1}}{2} \tag{9.1}$$

The size of $\mathcal{W}$ is therefore $|\mathcal{W}| = 8119$.

Given two words $w, w' \in \mathcal{W}$, we define $\mathcal{D}_p^{w,w'}$ as the family of subsets D of P_p such that:

- D dominates $I(P_p)$,
- w is the input word $w^{in}(D)$,
- w' is the output word $w_p^{out}(D)$.

A relevant information for our calculation will be to know, for two given words $w, w' \in \mathcal{W}$, the minimum loss over all losses $\ell(D)$ where $D \in \mathcal{D}_p^{w,w'}$. For this purpose, we introduce the 8119×8119 square matrix C_p. For $w, w' \in \mathcal{W}$, let $C_p[w, w'] = \min_{D \in \mathcal{D}_p^{w,w'}} \ell(D)$ where the minimum of the empty set is $+\infty$.

Let $w, w' \in \mathcal{W}$ be two given words. Due to the symmetry of P_{12} with respect to the first diagonal (bottom-left to top-right) of the grid, if a vertex set D belongs to $\mathcal{D}_{12}^{w,w'}$, then $D' = \{(j,i) | (i,j) \in D\}$ belongs to $\mathcal{D}_{12}^{w',w}$. Moreover, it is clear that, always due to the symmetry, $\ell(D) = \ell(D')$. Therefore, we have $C_{12}[w, w'] = C_{12}[w', w]$ and thus C_{12} is a symmetric matrix. Despite the size of C_{12} and the size of P_{12} (141 vertices), it is possible to compute C_{12} in less than one hour by computer. Indeed, we choose a sequence of subsets $X_0 = \emptyset, X_1, \ldots, X_{141} = P_{12}$ such that for every $i \in \{1, \ldots, 141\}$, $X_i \subseteq X_{i+1}$ and $X_{i+1} \setminus X_i = \{x_{i+1}\}$. Moreover, we choose the sequence such that for every i, $X_i \setminus I(X_i)$ is at most 21. This can be done for example by taking $x_{i+1} = \min\{(x,y) : (x,y) \in P_{12} \setminus X_i\}$, where the order is the lexical order. For $i \in \{0, \ldots, 141\}$, we compute for every labeling $f \in \mathcal{F}_i$, where $\mathcal{F}_i$ is the set of functions $(X_i \setminus I(X_i)) \to \{0,1,2\}$, the minimal loss $l_{i,f}$ of a set $S \subseteq X_i$ which dominates $I(X_i)$ and such that $\phi_S(v) = f(v)$ for every $v \in X_i \setminus I(X_i)$. Note that not every labeling is possible since two adjacent vertices cannot be labeled 0 and 2. The number of possible labellings can be computed using formula (9.1), and since $|X_i \setminus I(X_i)|$ can be covered by a path of at most 23 vertices, this gives, in the worst case, that this number is less than 10^9 and can be then processed by a computer. We compute inductively the sequence $(l_{i,f})_{f \in \mathcal{F}_i}$ from the sequence $(l_{i-1,f})_{f \in \mathcal{F}_{i-1}}$ by dynamical programming, and C is easily deduced from $(l_{141,f})_{f \in \mathcal{F}_{141}}$.

In the following, our aim is to glue $P_{n-12}, Q_{m-12}, R_{n-12}$, and O_{m-12} together. For two consecutive parts of the border, say P_{n-12} and Q_{m-12}, the output word of Q_{m-12} should be compatible with the input word of P_{n-12}. Two words w, w' of $\mathcal{W}$ are *compatible* if the sum of their corresponding letters is at most 2, i.e. $w[i] + w'[i] \leq 2$ for all $i \in [9]$. Note that $w[10] + w'[10]$ should be greater than 2 since the corresponding vertices can be dominated by some vertices of $V_{n,m} \setminus B_{n,m}$.

Given two words $w, w' \in \mathcal{W}$, let $\ell(w, w') = |\{i \in [10] : w[i] \neq 2 \text{ and } w'[i] = 0\}| + |\{i \in [10] : w'[i] \neq 2 \text{ and } w[i] = 0\}|$.

Lemma 9.7. *Let D be a dominating set of $G_{n,m}$ and let us denote $D_P = D \cap P_{n-12}$ and $D_Q = D \cap Q_{m-12}$. Then $\ell(D \cap (P_{n-12} \cup Q_{m-12})) = \ell(D_P) + \ell(D_Q) + \ell(w, w')$, where $w = w^{in}(D_P)$ and $w' = w_q^{out}(f_{n,m}^{-1}(D_Q))$. Moreover, w and w' are compatible.*

Proof. By Proposition 9.5(ii), $\ell(D \cap (P_{n-12} \cup Q_{m-12})) = \ell(D_P) + \ell(D_Q) + |N[D_P] \cap N[D_Q]|$. It suffice then to note that $\ell(w, w') = |N[D_P] \cap N[D_Q]|$ to get $\ell(D \cap (P_{n-12} \cup Q_{m-12})) = \ell(D_P) + \ell(D_Q) + \ell(w, w')$.

In what remains, we prove that w and w' are compatible. If those two words were not compatible, there would exist an index $i \in [9]$ such that

$$w_{m-12}^{out}(f_{n,m}^{-1}(D_Q))[i] + w^{in}(D_P)[i] > 2.$$

Thus at least one of these two letters should be a 2, and the other one should not be 0.

Suppose that $w_{m-12}^{out}(f_{n,m}^{-1}(D_Q))[i] = 2$ and note that this means that the vertex $(i, 13)$ is not dominated by a vertex in D_Q. Since D is a dominating set of $G_{n,m}$, every output vertex of Q_{m-12} except $(10, 13)$ (and every input vertex of P_{n-12} except $(10, 12)$) is dominated by a vertex of D_Q or by a vertex of D_P. Thus $(i, 13)$ should be dominated by its unique neighbour in P_{n-12}, $(i, 12)$. This would imply that $(i, 12) \in D$ contradicting the fact that $w^{in}(D_P)[i] \ne 0$.

Similarly, if $w^{in}(D_P)[i] = 2$, the vertex $(i, 12)$ is not dominated by a vertex in D_P, thus $(i, 12)$ must be dominated by the vertex $(i, 13) \in D$, contradicting the fact that $w_{m-12}^{out}(f_{n,m}^{-1}(D_Q))[i] \ne 0$. $\square$

Lemma 9.7 is designed for the two consecutive parts P_{n-12} and Q_{m-12} of the border of $G_{n,m}$. Its easy to see that this extends to any pair of consecutive parts of the border, i.e. Q_{m-12} and R_{n-12}, R_{n-12} and O_{m-12}, O_{m-12} and P_{n-12}.

We define the matrix 8119×8119 square matrix L which contains, for every pair of words $w, w' \in \mathcal{W}$, the value $\ell(w, w')$:

$$L[w, w'] = \begin{cases} +\infty \text{ if } w \text{ and } w' \text{ are not compatible,} \\ \ell(w, w') \text{ otherwise.} \end{cases}$$

Note that L is symmetric since $\ell(w, w') = \ell(w', w)$.

Let $\otimes$ be the matrix multiplication in $(\min, +)$ algebra, i.e. $C = A \otimes B$ is the matrix where for all i, j, $C[i, j] = \min_k A[i, k] + B[k, j]$.

Let $M_p = L \otimes C_p$ for $p \ge 12$.

By construction, $M_{n-12}[w, w']$ corresponds to the minimum possible loss $\ell(D \cap P_{n-12})$ of a dominating set $D \subseteq V_{n,m}$ that dominates $I(P_{n-12})$ and such that w is the output word of Q_{m-12} and w' is the output word of P_{n-12}.

Lemma 9.8. *For all $24 \le n \le m$, we have*

$$b_{n,m} \ge \min_{\substack{w_1 \in \mathcal{W} \\ w_2 \in \mathcal{W} \\ w_3 \in \mathcal{W} \\ w_4 \in \mathcal{W}}} M_{n-12}[w_1, w_2] + M_{m-12}[w_2, w_3] + M_{n-12}[w_3, w_4] + M_{m-12}[w_4, w_1].$$

Proof. Consider a set $D \subseteq B_{n,m}$ which dominates $I(B_{n,m})$ and achieving $\ell(D) = b_{n,m}$. Let $D_P = D \cap P_{n-12}$, $D_Q = D \cap Q_{m-12}$, $D_R = D \cap R_{n-12}$ and $D_O =$

$D \cap O_{m-12}$. Let w_P (w_Q, w_R and w_O, respectively) be the input word of P_{n-12} (Q_{m-12}, R_{n-12} and O_{m-12}), and w'_P (w'_Q, w'_R and w'_O) be the output word of P_{n-12} (Q_{m-12}, R_{n-12} and O_{m-12}). By definition of C_p, the loss of D_P is at least $C_{n-12}[w_P, w'_P]$. Similarly, we have $\ell(D_Q) \geq C_{m-12}[w_Q, w'_Q]$, $\ell(D_R) \geq C_{n-12}[w_R, w'_R]$ and $\ell(D_O) \geq C_{m-12}[w_O, w'_O]$. By the definition of the loss:

$$\begin{aligned}
\ell(D) &= b_{n,m} \\
&= 5 \times |D| - |N[D]| \\
&= \ell(D_O) + \ell(D_P) + \ell(D_Q) + \ell(D_R) \\
&\quad + L[w'_O, w_P] + L[w'_P, w_Q] + L[w'_Q, w_R] + L[w'_R, w_O] \\
&\qquad \text{by Lemma 9.7 and since } N[D_P] \cap N[D_R] = N[D_Q] \cap N[D_O] = \emptyset \\
&\geq C_{m-12}[w_O, w'_O] + C_{n-12}[w_P, w'_P] + C_{m-12}[w_Q, w'_Q] + C_{n-12}[w_R, w'_R] \\
&\quad + L[w'_O, w_P] + L[w'_P, w_Q] + L[w'_Q, w_R] + L[w'_R, w_O] \\
&\geq M_{m-12}[w_O, w_P] + M_{n-12}[w_P, w_Q] + M_{m-12}[w_Q, w_R] + M_{n-12}[w_R, w_O] \\
&\qquad \text{since } w'_O \text{ and } w_P \text{ (resp. } w'_P \text{ and } w_Q,\ w'_Q \text{ and } w_R,\ w'_R \text{ and } w_O) \\
&\qquad \text{are compatibles.}
\end{aligned}$$

□

According to Lemma 9.8, to bound $b_{n,m}$ it would be thus interesting to know M_p for $p > 12$. It is why we introduce the following 8119×8119 square matrix, T.

Lemma 9.9. *There exists a matrix T such that $C_{p+1} = C_p \otimes T$ for all $p \geq 12$. This matrix is defined as follows:*

$$T[w, w'] = \begin{cases} +\infty & \text{if } \exists i \in [10] \text{ s.t. } w[i] = 0 \text{ and } w'[i] = 2 \\ +\infty & \text{if } \exists i \in [9] \text{ s.t. } w[i] = 2 \text{ and } w'[i] \neq 0 \\ +\infty & \text{if } \exists i \in \{2, \ldots, 9\} \text{ s.t. } w'[i] = 1,\ w[i] \neq 0, \\ & \quad w'[i-1] \neq 0 \text{ and } w'[i+1] \neq 0 \\ +\infty & \text{if } w'[1] = 1,\ w[1] \neq 0 \text{ and } w'[2] \neq 0 \\ +\infty & \text{if } w'[10] = 1,\ w[10] \neq 0 \text{ and } w'[9] \neq 0 \\ 3 \times |w'|_0 - |w|_2 - |w'|_1 + |w|_0 - 1 & \text{if } w'[10] = 0 \\ 3 \times |w'|_0 - |w|_2 - |w'|_1 + |w|_0 & \text{otherwise.} \end{cases}$$

Proof. Consider a set $S' \subseteq P_{p+1}$ dominating $I(P_{p+1})$ and let $S = S' \cap P_p$. Let $w = w_p^{out}(S)$ and $w' = w_{p+1}^{out}(S')$. Let $\Delta(S, S') = \ell(S') - \ell(S)$. By definition of the loss, $\Delta(S, S') = 5 \times |S' \backslash S| - |N[S'] \backslash N[S]|$. Let us compute $\Delta(S, S')$ in term of the number of occurrences of 0's, 1's and 2's in the words w and w'. The set $S' \backslash S$ corresponds to the vertices $\{(p+1, i) \mid i \in [10], w'[i] = 0\}$. The set $N[S'] \setminus N[S]$ corresponds to the vertices dominated by S' but not dominated by S; these vertices clearly belong to the columns p, $p+1$ and $p+2$. Since S' dominates $I(P_{p+1})$, those in the column p are the vertices $\{(p, i) \mid i \in [10], w[i] = 2\}$. Those in the column $p+1$ are the vertices $\{(p+1, i) \mid i \in [10], w'[i] \neq 2, w[i] \neq 0\}$ and possibly the vertex $(p+1, 11)$ when $w'[10] = 0$. Finally, those in the column

$p+2$ are the vertices $\{(p+2,i) \mid i \in [10], w'[i]=0\}$. We then get:

$$\Delta(S,S') = \begin{cases} 3 \times |w'|_0 - |w|_2 - |w'|_1 + |w|_0 - 1 & \text{if } w'[10]=0 \\ 3 \times |w'|_0 - |w|_2 - |w'|_1 + |w|_0 & \text{otherwise} \end{cases}$$

where $|w|_n$ denotes the number of occurrences of the letter n in the word w.

Thus $\Delta(S,S')$ only depends on the output words of S and S', and we can denote this value by $\Delta(w,w')$. Note however that there exist pairs of words (w,w') that could not be the output words of S and S'; there are three cases:

Case 1. $w[i]=0$ and $w'[i]=2$ since the vertex $(p+1,i)$ would be dominated by (p,i) contradicting its label 2;

Case 2. $w[i]=2$ and $w'[i] \neq 0$ for $i \in [9]$ since (p,i) would not be dominated contradict the fact that S' dominates $I(P_{p+1})$;

Case 3. $w'[i]=1$ and $w'[i-1] \neq 0$, $w'[i+1] \neq 0$, $w[i] \neq 0$ since $(p+1,i)$ would be dominated according to its label but none of its neighbors belong to S'.

For these forbidden cases, we set $\Delta(w,w') = +\infty$.

By definition, $C_{p+1}[w_i,w']$ is the minimum loss $\ell(S')$ of a set $S' \subseteq P_{p+1}$ that dominates $I(P_{p+1})$, with w_i as input word and w' as output word. It is clear that $S = S' \cap P_p$ dominates $I(P_p)$ and has w_i as input word. Let w be its output word and note that $C_{p+1}[w_i,w'] = \ell(S') = \ell(S) + \Delta(w_i,w')$. The minimality of $\ell(S')$ implies the minimality of $\ell(S)$ over the sets $X \in \mathcal{D}_p^{w_i,w'}$. Indeed, any set $X \in \mathcal{D}_p^{w_i,w'}$ could be turned in a set of $X' \in \mathcal{D}_{p+1}^{w_i,w'}$ by adding vertices of the $p+1^{th}$ column accordingly to w'. Thus

$$C_{p+1}[w_i,w'] = C_p[w_i,w] + \Delta(w,w')$$

which implies that

$$C_{p+1}[w_i,w'] \geq \min_w C_p[w_i,w] + \Delta(w,w').$$

On the other hand, for every word $w_o \in \mathcal{W}$ such that $C_p[w_i,w_o] \neq +\infty$ and $\Delta(w_o,w') \neq +\infty$, there is a set $S \in \mathcal{D}_p^{w_i,w_o}$, with $\ell(S) = C_p[w_i,w_o]$, that can be turned in a set $S' \in D_{p+1}^{w_i,w'}$ with $\ell(S') = C_p[w_i,w_o] + \Delta(w_o,w')$. Thus

$$C_{p+1}[w_i,w'] \leq \min_{w_o} C_p[w_i,w_o] + \Delta(w_o,w').$$

This concludes the proof of the lemma. □

By the definition of M_p, we have also $M_{p+1} = M_p \otimes T$. Note that T is a sparse matrix: about 95.5% of its 8119^2 entries are $+\infty$. Thus the multiplication by T in the $(\min,+)$ algebra can be done in a reasonable amount of time by a trivial algorithm.

Fact 9.10. The computations give us that $M_{126} = M_{125} + 1$. Thus, since $(A+c) \otimes B = (A \otimes B) + c$ for any matrices A, B and any integer c, we have that $M_{125+k} = M_{125} + k$ for every $k \in \mathbb{N}$.

Let us define $M'_p = \min_{k\in\mathbb{N}}(M_{p+k}-k)$. Then, for all $q \geq p$, $M_q \geq M'_p+(q-p)$. By Fact 9.10, $M'_p = \min_{k\in\{0,\dots 125-p\}}(M_{p+k}-k)$

Fact 9.11. By computing M'_{12}, and $A' = M'_{12}\otimes M'_{12}$, we obtain that $\min_{w_1,w_3}(A'+A'^T)[w_1,w_3] = 76$ (where A^T is the transpose of A).

This implies that

$$\begin{aligned}
\min_{w_1,w_3}\ (\min_{w_2} M'_{12}[w_1,w_2]+M'_{12}[w_2,w_3]) + (\min_{w_4} M'_{12}[w_3,w_4]+M'_{12}[w_4,w_1]) &= 76\\
\min_{w_1,w_2,w_3,w_4} M'_{12}[w_1,w_2]+M'_{12}[w_2,w_3]+M'_{12}[w_3,w_4]+M'_{12}[w_4,w_1] &= 76.
\end{aligned}$$

Theorem 9.12. *If $24 \leq n \leq m$, then*

$$\gamma(G_{n,m}) = \left\lfloor \frac{(n+2)(m+2)}{5} \right\rfloor - 4.$$

Proof. By Chang's construction [56], $\gamma(G_{n,m}) \leq \left\lfloor \frac{(n+2)(m+2)}{5} \right\rfloor - 4$. Let us now compute a lower bound for the loss of a dominating set of $G_{n,m}$.

$$\begin{aligned}
\ell_{n,m} &\geq b_{n,m}\\
&\geq \min_{w_1,w_2,w_3,w_4} M_{n-12}[w_1,w_2]+M_{m-12}[w_2,w_3]\\
&\qquad\qquad + M_{n-12}[w_3,w_4]+M_{m-12}[w_4,w_1]\\
&\quad \text{by Lemma 9.8}\\
&\geq \min_{w_1,w_2,w_3,w_4} M'_{12}[w_1,w_2]+(n-12-12)\\
&\qquad\qquad + M'_{12}[w_2,w_3]+(m-12-12)+M'_{12}[w_3,w_4]\\
&\qquad\qquad + (n-12-12)+M'_{12}[w_4,w_1]+(m-12-12)\\
&\geq 2\times(n+m-48)\\
&\quad + \min_{w_1,w_2,w_3,w_4} M'_{12}[w_1,w_2]+M'_{12}[w_2,w_3]+M'_{12}[w_3,w_4]+M'_{12}[w_4,w_1]\\
&\geq 2\times(n+m-48)+76\\
&\geq 2\times(n+m)-20.
\end{aligned}$$

Thus by Lemma 9.6, we have:

$$\begin{aligned}
\gamma(G_{n,m}) &\geq \left\lceil \frac{n\times m+2\times(n+m)-20}{5} \right\rceil\\
&\geq \left\lceil \frac{(n+2)(m+2)-4}{5} \right\rceil - 4\\
&\geq \left\lfloor \frac{(n+2)(m+2)}{5} \right\rfloor - 4.
\end{aligned}$$

$\square$

Chapter 10

An aperiodic set of 11 Wang tiles

This chapter presents a joint work with Emmanuel Jeandel [11]. A new aperiodic tile set containing 11 Wang tiles on 4 colors is presented. This tile set is minimal in the sense that no Wang set with less than 11 tiles is aperiodic, and no Wang set with less than 4 colors is aperiodic.

Introduction. Wang tiles are square tiles with colored edges. A tiling of the plane by Wang tiles consists in putting a Wang tile in each cell of the grid $\mathbb{Z}$ so that contiguous edges share the same color. The formalism of Wang tiles was introduced by Wang [147] to study decision procedures for a specific fragment of logic (see section 10.1.1 for details).

Wang asked the question of the existence of an aperiodic tile set: A set of Wang tiles which tiles the plane but cannot do so periodically. His student Berger quickly gave an example of such a tile set, with a tremendous number of tiles. The number of tiles needed for an aperiodic tileset was reduced during the years, first by Berger himself, then by others, to obtain in 1996 the previous record of an aperiodic set of 13 Wang tiles. (see section 10.1.2 for an overview of previous aperiodic sets of Wang tiles).

While reducing the number of tiles may seem like a tedious exercise in itself, the articles also introduced different techniques to build aperiodic tilesets, and different techniques to prove aperiodicity.

A few lower bounds exist on the number of Wang tiles needed to obtain an aperiodic tile set, the only reference [81] citing the impossibility to have one with 4 tiles or less. On the other hand, recent results show that an aperiodic set of Wang tiles need to have at least 4 different colors [55].

In this article, we fill all the gaps: We prove that there are no aperiodic tile set with less than 11 Wang tiles, and that there is an aperiodic tile set with 11 Wang tiles and 4 colors.

The discovery of this tile set, and the proof that there is no aperiodic tile set with a smaller number of tiles was done by a computer search: We generated in particular all possible candidates with 10 tiles or less, and prove they were not aperiodic. Surprisingly it was somewhat easy to do so for all of them except one. The situation is different for 11 tiles: While we have found an aperiodic tileset, we also have a short list of tile set for which we do not know anything. The description of this computer search is described in section 10.3 of the paper, and can possibly be skipped by a reader only interested in the tile set itself. This section also contains a result of independent interest: the tile set from Culik with one tile omitted does not tile the plane.

The tile set itself is presented in section 10.4, and the remaining sections prove that it is indeed an aperiodic tileset.

10.1 Aperiodic sets of Wang tiles

Here is a brief summary of the known aperiodic sets of Wang tiles. Explanations about some of them may be found in [81]. We stay clear in this history about aperiodic sets of geometric figures, and focus only on Wang tiles.

10.1.1 Wang tiles and the $\forall\exists\forall$ problem

Wang tiles were introduced by Wang [147] in 1961 to study the decidability of the $\forall\exists\forall$ fragment of first order logic. Wang showed in this article how to build, starting from a $\forall\exists\forall$ formula ϕ, a set of tiles τ and a subset $\tau' \subseteq \tau$ so that there exists a tiling by τ of the upper quadrant with tiles in the first row in τ' iff ϕ is satisfiable. If this particular tiling problem was decidable, this would imply that the satisfiability of $\forall\exists\forall$ formulas was decidable.

Wang asked more generally in this article whether the more general tiling problem (with no particular tiles in the first row) is decidable and gave the *fundamental conjecture*: Every tileset either admits a periodic tiling or does not tile.

Regardless of the status of this particular conjecture, Kahr, Moore and Wang [95] proved the next year that the $\forall\exists\forall$ problem is indeed undecidable by reducing to another tiling problem: now we fix a subset τ' of tiles so that every tile on the diagonal of the first quadrant is in τ'. This proof was later simplified by Hermes [84, 85]. From the point of view of first order logic, the problem is thus solved. Formally speaking, the tiling problem with a constraint diagonal is reduced to a formula of the form $\forall x \exists y \forall z \phi(x, y, z)$ where ϕ contains a binary predicate P and some occurences of the subformula $P(x, x)$ (to code the diagonal constraint). If we look at $\forall\exists\forall$ formulas that do not contain the subformula $P(x, x)$ and $P(z, z)$, the decidability of this particular fragment remained open.

A few years later, Berger proved however [40] that the domino problem is undecidable, and that an aperiodic tileset existed. This implies in particular that the particular fragment of $\forall x \exists y \forall z$ where the only occurences of the binary predicates P are of the form $P(x, z), P(y, z), P(z, y), P(z, x)$ was undecidable.

A few other subcases of $\forall\exists\forall$ were done over the years. In 1975, Aanderaa and Lewis [33] proved the undecidability of the fragment of $\forall\exists\forall$ where the binary predicates P can only appear in the form $P(x, z)$ and $P(z, y)$. It has in particular the following consequence: The domino problem for *deterministic* tilesets is undecidable

10.1.2 Aperiodic tilesets

The first example of a set of Wang tiles was provided by Berger in 1964. The set contained in the 1966 AMS publication [41] contains 20426 tiles, but Berger's original PhD Thesis [40] also contains a simplified version with 104 tiles. This tileset is of a substitutive nature. Knuth [107] gave another simplified version of Berger's original proof with 92 tiles.

Lauchli obtained in 1966 an aperiodic set of 40 Wang tiles, published in 1975 in a paper of Wang [146].

Robinson found in 1967 an aperiodic set of 52 tiles. It was mentioned in a Notices of the AMS summary, but the only place this set can be found is in an article of Poizat [128]. His most well known tileset is however a 1969 tileset (published in 1971) [134] of 56 tiles. The paper hints at a set of 35 Wang tiles.

Robinson managed to lower the number of tiles again to 32 using a idea due to Roger Penrose. The same idea is used by Grunbaum and Shephard to obtain an aperiodic set of 24 tiles [81]. Robinson obtained in 1977 a set of 24 tiles from a tiling method by Ammann. The record for a long time was held by Ammann, who obtained in 1978 a set of 16 Wang tiles. Details on these tilesets are provided when available in [81]

In 1975, Aanderaa and Lewis [33] build the first aperiodic *deterministic* tileset. No details about the tileset are provided but it is possible to extract one from the exposition by Lewis [110]. This construction was somehow forgotten in the literature and the first aperiodic deterministic tileset is usually attributed to Kari in 1992 [101].

In 1989, Mozes showed a general method that can be used to translate any substitution tiling into a set of Wang tiles [119], which will be of course aperiodic. There are multiple generalizations of this result (depending of the exact definition of "substitution tiling"), of which we cite only a few [80, 74, 79]. For a specific example, Socolar build such a representation [137] of the chair tiling, which in our vocabulary can be done using 64 tiles.

The story stopped until 1996 when Kari invented a new method to build aperiodic tileset and obtained an aperiodic set of 14 tiles [100]. This was reduced to 13 tiles by Culik [58] using the same method. There was suspicion one of the 13 tiles was unnecessary, and Kari and Culik hinted to a method to show it in a unpublished manuscript. However this is not true: the method developed in this article will show this is not the case.

In 1999, Kari and Papasoglu [102] presented the first 4-way deterministic aperiodic set. The construction was later adapted by Lukkarilla to provided a proof of undecidability of the 4-way domino problem [113].

The construction of Robinson was later analyzed [136, 35, 93, 78] and simplified. Durand, Levin and Shen presented in 2004 [66] a way to simplify exposition of proofs of aperiodicity of such tilesets. Ollinger used this method in 2008 to obtain an aperiodic tileset with 104 tiles [123], with striking resemblance to the original set of 104 tiles by Berger. Other simplifications of Robinson constructions where given by Levin in 2005 [109] and Poupet in 2010 [129] using ideas similar to Robinson.

In 2008, Durand, Romashchenko and Shen provided a new construction based on the classical fixed point construction from computability theory [68, 67].

10.2 Preliminaries

10.2.1 Wang tiles

A *Wang tile* is a unit square with colored edges. Formally, let H, V be two finite sets (the horizontal and vertical colors, respectively). A wang tile t is an element of $H^2 \times V^2$. We write $t = (t_w, t_e, t_s, t_n)$ for a Wang tile, and use interchangeably the notations t_w (resp. t_e, t_s, t_n) or $w(t)$ (resp. $e(t), s(t), n(t)$) to indicate the color on one of the edges.

A *Wang set* is a set of Wang tiles, formally viewed as a tuple (H, V, T), where $T \subseteq H^2 \times V^2$ is the set of *tiles*. Fig. 10.1 presents a well known example of a Wang set. A Wang set is said to be *empty* if $T = \emptyset$.

Let $\mathcal{T} = (H, V, T)$ be a Wang set. Let $X \subseteq \mathbb{Z}^2$. A *tiling of X by $\mathcal{T}$* is an assignation of tiles from $\mathcal{T}$ to X so that contiguous edges have the same color, that is it is a function $f : X \to T$ such that $e(f(x, y)) = w(f(x + 1, y))$ and $n(f(x, y)) = s(f(x, y + 1))$ for every $(x, y) \in \mathbb{Z}^2$ when the function is defined. We are especially interested in the tilings of $\mathbb{Z}^2$ by a Wang set $\mathcal{T}$. When we say a *tiling of the plane by* $\mathcal{T}$, or simply a *tiling by* $\mathcal{T}$, we mean a tiling of $\mathbb{Z}^2$ by $\mathcal{T}$.

A tiling f is *periodic* if there is a $(u, v) \in \mathbb{Z}^2 \setminus (0,0)$ such that $f(x, y) = f(x+u, y+v)$ for every $(x, y) \in \mathbb{Z}^2$. A tiling is *aperiodic* if it is not periodic.

A Wang set *tiles* X (resp. *tiles the plane*) if there exists a tiling of X (resp. the plane) by $\mathcal{T}$. A Wang set is *finite* if there is no tiling of the plane by $\mathcal{T}$. A Wang set is *periodic* if there is a tiling t by $\mathcal{T}$ which is periodic. A Wang set is *aperiodic* if it tiles the plane, and every tiling by $\mathcal{T}$ is not periodic.

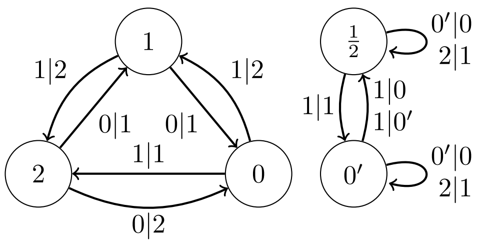
We quote here a few well known folklore results:

Lemma 10.1. *If $\mathcal{T}$ is periodic, then there is a tiling t by $\mathcal{T}$ with two linearly independent translation vectors (in particular a tiling t with vertical and horizontal translation vectors).*

Lemma 10.2. *If for every $k \in \mathbb{N}$, there exists a tiling of $[0, \ldots, k] \times [0, \ldots, k]$ by $\mathcal{T}$, then $\mathcal{T}$ tiles the plane.*

10.2.2 Transducers

One of the most trivial but crucial observation we will use in this article is that Wang sets (H, V, T) may be viewed as finite state transducers, where each transition reads and writes one letter, and without initial nor final states: H is the set of states, V is the input and output alphabet, and T is the set of transitions. Fig. 10.1 presents in particular the popular set of Wang tiles introduced by Culik from both point of views.



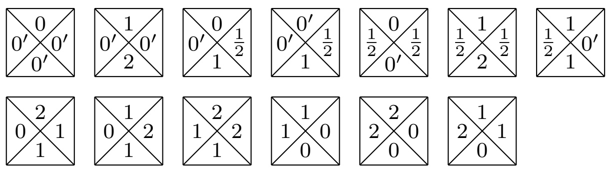


Figure 10.1 – The aperiodic set of 13 tiles obtained by Culik from an idea of Kari: the transducer view and the tiles view.

In this formalism, tilings correspond exactly to (biinfinite) runs of the transducer. If w and w' are biinfinite words over the alphabet V, we will write $w\mathcal{T}w'$ if w' is the image of w by the transducer. The transducer is usually nondeterministic so that this is indeed a (partial) relation and not a function.

The composition of Wang sets, seen as transducers, is straightforward: Let $\mathcal{T} = (H, V, T)$ and $\mathcal{T}' = (H, V', T')$ be two Wang sets. Then $\mathcal{T} \circ \mathcal{T}'$ is the Wang set $(H \times H', V, T'')$, where

$$T'' = \{((w, w'), (e, e'), s, n') : (w, e, s, n) \in T, (w', e', s', n') \in T' \text{ and } n = s'\}.$$

Let $\mathcal{T}^k$, $k \in \mathbb{N}^*$ be $\mathcal{T}$ if $k = 1$, $\mathcal{T}^{k-1} \circ \mathcal{T}$ otherwise.

A reformulation of the original question is as follows:

Lemma 10.3. *A Wang set $\mathcal{T}$ is finite if there is no infinite run of the transducer $\mathcal{T}$: There is no biinfinite sequence $(w_k)_{k\in\mathbb{N}}$ so that $w_k \mathcal{T} w_{k+1}$ for all k.*

A Wang set $\mathcal{T}$ is periodic iff there exists a word w and a positive integer k so that $w\mathcal{T}^k w$.

We will also use the following operations on tile sets (or transducers):

rotation Let $\mathcal{T}^{\mathrm{tr}}$ be (V, H, T') where $T' = \{(s, n, e, w) : (w, e, s, n) \in T\}$. This operation corresponds to a rotation of the tileset by 90 degrees.

simplification Let $s(\mathcal{T})$ be the operation that deletes from $\mathcal{T}$ any tile that cannot be used in a tiling of a (biinfinite) line row by $\mathcal{T}$. This corresponds from the point of view of transducers to eliminating sources and sinks from $\mathcal{T}$. In particular $s(\mathcal{T})$ is empty iff there is no words w, w' s.t. $w\mathcal{T}w'$.

union $\mathcal{T} \cup \mathcal{T}'$ is the disjoint union of transducers $\mathcal{T}$ and $\mathcal{T}'$: We first rename the states of both transducers so that they are all different, and then we take the union of the transitions of both transducers. Thus $w(\mathcal{T} \cup \mathcal{T}')w'$ iff $w\mathcal{T}w'$ or $w\mathcal{T}'w'$.

Equivalence of Wang sets. Once Wang sets are seen as transducers, it is easy to see that the problems under consideration do not depend actually on $\mathcal{T}$, but only on the relation induced by $\mathcal{T}$: We say that two Wang sets $\mathcal{T} = (H, V, T)$ and $\mathcal{T}' = (H', V, T')$ are *equivalent* if they are equivalent as relations, that is, for every pair of bi-infinite words (w, w') over V, $w\mathcal{T}w' \Leftrightarrow w\mathcal{T}'w'$.

In the course of the proofs and the algorithms, it will be interesting to switch between equivalent Wang sets (transducers), in particular by trying to simplify as much as possible the sets: we can for example apply the operator $s(\mathcal{T})$ to trim the colors/states (and thus the tiles/transitions) that cannot appear in a infinite row (e.g. sources/terminals of the transducer seen as a graph), or reduce the size of the transducer by coalescing "equivalent" states.

There are a few algorithms to simplify Wang sets. First, as our transducers are nothing but (nondeterministic) finite automata over the alphabet $V \times V$, it is tempting to try to *minimise* them. However state (or transition) minimisation of nondeterministic automata is PSPACE-complete ; The other strategy of building the minimal determinic automaton is also not efficient in practice. The algorithm we used is based on the notion of *strong bisimulation equivalence* of labeled transitions sytems [96, 124, 145, 112]. It allows us to find efficiently states that are equivalent (in some sense) and thus can be collapsed together. It can be thought of as the non-deterministic equivalent of the classical minimization algorithm for deterministic automata from Hopcroft [86].

10.3 There is no aperiodic Wang sets with 10 tiles or less

In this section, we give a brief overview of the techniques involved in the computer assisted proof that there are not aperiodic Wang set with 10 tiles or less.

The general method of the algorithm is obvious: generate all Wang sets with 10 tiles or less, and test whether there are aperiodic. There are two difficulties here: first, there are a large number of Wang sets with 10 tiles: For maximum efficiency, we have to generate as few of them as possible, that is discard as soon as possible Wang sets that are provably not aperiodic. Then we have to test the remaining sets for aperiodicity. Aperiodicity is of course an undecidable problem: our algorithm will not succeed on all Wang sets, and the remaining ones will have to be examined by hand.

10.3.1 Generating all Wang sets with 10 tiles or less

According to the general principle above, we actually do not have to generate all Wang sets: we can refrain from generating sets that we know are not aperiodic.

Let $\mathcal{T}$ be a Wang set. We say that $\mathcal{T}$ is minimally aperiodic if $\mathcal{T}$ is aperiodic and no proper subset of $\mathcal{T}$ is aperiodic (that is no proper subset of $\mathcal{T}$ tiles the plane). We will introduce criteria proving that some Wang sets are not minimally aperiodic, and thus that we do not need to test them.

The key idea is to look at the graph G underlying the transducer. Note that this is actually a *multigraph*: there might be multiple edges(transitions) joining two given vertices (states), and there might also be self-loops.

This approach was also introduced in [92], and the following lemma is more or less implicit in this article:

Lemma 10.4. *Let $\mathcal{T}$ be a Wang set, and G the corresponding graph.*

- *Suppose there exist two vertices/states/colors $u, v \in G$ so that there is an edge (hence a tile/transition) from u to v and no path from v to u. Then $\mathcal{T}$ is not minimal aperiodic.*

- *Suppose G contains a strongly connected component which is reduced to a cycle. Then $\mathcal{T}$ is not minimal aperiodic.*

- *If the difference between the number of edges and the number of vertices in G is less than 2, then $\mathcal{T}$ is not minimal aperiodic.*

Proof. In terms of tiles, the first case corresponds to a tile t that can appear at most one in each row. If $\mathcal{T}$ tiles the plane, $\mathcal{T}$ tiles arbitrarily large regions without using the tile t. By compactness, $\mathcal{T} \setminus \{t\}$ tiles the plane.

For the second case, suppose such a component exists. This means there exist some tiles $S \subseteq \mathcal{T}$ so that every time one of the tiles in S appear, then the whole row is periodic (of period the size of the cycle). If $\mathcal{T}$ is aperiodic, we cannot have a tiling where tiles of S appear in two different rows, as we could deduce from it a periodic tiling. As a consequence, tiles from S appear in at most one row, and using the same compactness argument as before we deduce that $\mathcal{T} \setminus S$ tiles the plane.

The proof of the third case can be found in [92]. □

This lemma gives a bird's eye-view of the program: For a given $n \leq 10$, generate all (multi)graphs G with n edges and at most $n-2$ vertices satisfying the hypotheses of the lemma, then test all Wang sets for which the underlying graph in G. In terms of Wang tiles, a graph correspond to a specific assignation of colors to the east/west side: for this particular assignation, we test all possible assignations of colors to the north/south side.

The exact approach used in the software follows this principle, trying as much as possible not to generate isomorphic tilesets.

10.3.2 Testing Wang sets for aperiodicity

We explained in the previous section how we generated Wang sets to test. We now explain how we tested them for aperiodicity.

Easy cases

Recall that a Wang set is *not* aperiodic if

- Either there exists k so that $s(\mathcal{T}^k)$ is empty: there is no word w, w' so that $w\mathcal{T}^k w'$
- or there exists k so that $\mathcal{T}^k$ is periodic: there exists a word w so that $w\mathcal{T}^k w$

The general algorithm to test for aperiodicity is therefore clear: for each k, generate $\mathcal{T}^k$, and test if one of the two situations happen. If it does, the set is not aperiodic. Otherwise, we go to the next k. The algorithm stops when the computer program runs out of memory. In that case, the algorithm was not able to decide if the Wang set was aperiodic (it is after all an undecidable problem), and we have to examine carefully this Wang set.

This approach works quite well in practice: when launched on a computer with a reasonable amount of memory, it eliminates a very large number of tilesets. Of course, the key idea is to simplify as much as possible $\mathcal{T}^k$ before computing $\mathcal{T}^{k+1}$. Note that this should be done as fast as possible, as this will be done for *all* Wang sets. It turns out that the easy simplification that consists in deleting at each step tiles that cannot appear in a tiling of a row (i.e. vertices that are sources/terminals) is already sufficient.

It is important to note that this approach relying on transducers (test whether the Wang set tiles k consecutive rows, and if it does so periodically) turned out in practice to be much more efficient than the naive approach using tilings of squares (test whether the Wang set tiles a square of size k, and if it does so periodically).

Harder cases

Once this has been done, a small number of Wang sets remain (at most 200), for which the program was not able to prove that they tile the plane periodically or that they do not tile the plane.

The first step for these sets was to use the same idea as before, but with a larger memory, and additional optimizations, which involved simplifying $\mathcal{T}^k$ as much as possible before examining it. Two additional simplifications were used: First, we may delete from $\mathcal{T}^k$ tiles(transitions) that connect different strongly connected components: using the same argument as in Lemma 10.4, it is easy to see that deleting these tiles do not change the aperiodic status of $\mathcal{T}^k$. Second, we have applied bisimulation techniques to reduce as much as possible the size

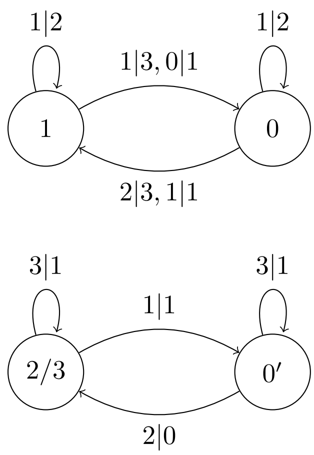


Figure 10.2 – A set of 10 tiles that tries very hard, but fails to tile the plane. It tiles however a square of size 212×212

of the transducer $\mathcal{T}^k$. We want to stress that this technique proved to be crucial: the gain obtained by bisimulation is tremendous.

The hardest case

Using these ameliorations on the ~ 200 remaining Wang sets was successful: all of them were quickly proven to be not aperiodic. Well not entirely! One small set of indomitable tiles still held out against the program.

This particular set of tiles is presented in Fig. 10.2. It turned out that this particular Wang set is a special case of a general construction introduced by Kari [100] of aperiodic Wang sets, except a few tiles are missing. At this point, the situation could have become desperate: It is not known if tilings obtained by the method of Kari but missing a few tiles may tile the plane. In fact, it was open whether it was possible to delete a tile from the 13 tileset from Culik [58] to obtain a set that still tiles the plane[1] (and it was conjectured by both Kari and Culik that it was indeed possible).

However we were able to prove that this tileset does not in fact tile the plane. Wang sets belonging to the family identified by Kari all work in the same way: The infinite words that appear on each row can be thought of as reals, by taking the average of all numbers (between 0 and 3 in our example) that appear on the row. Then what the tileset is doing is applying a given piecewise affine map to the real number. In the case of our set of 10 tiles, the map f is as follows:

- If $1/2 \leq x \leq 3/2$, then $f(x) = 2x$
- If $3/2 \leq x \leq 3$, then $f(x) = x/3$

[1]You will find many experts on tilings that recollect this story wrongly and think that the (13) Wang set by Culik is the (14) Wang set from Kari with one tile removed. It is not the case. What happened is that there is one tile from the (13) Wang set by Culik that seemed likely to be unnecessary.

As can be seen from the first transducer, there cannot be two consecutive 0 in x, this guarantees that $x \geq 1/2$ hence $x \neq 0$, and in particular that this tileset has no periodic tiling.

If we used the general method by Kari to code this particular tileset, the transducer that divides by 3 would have 8 tiles. However, our particular set of 10 tiles does so with only 4 tiles. There is a way to explain how the division by 3 works. First, let's see it like a multiplication by 3 by reversing the process. Recall that the Beatty expansion of a real x is given by $\beta_n(x) = \lceil (n+1)x \rceil - \lceil nx \rceil$. Then it can be proven:

Fact 10.5. Let $0 < x \leq 1$ and define $b_n(x) = 2\beta_n(2x) - \beta_n(x)$. Then the second transducer transforms $(\beta_n)_{n\in\mathbb{N}}$ into $(b_n)_{n\in\mathbb{N}}$.

Hence, the second transducer multiplies by 3 by doing $2 \times 2 \times x - x$ somehow. It can be seen as a composition of a transducer that transforms $(\beta_n)_{n\in\mathbb{N}}$ into $(\beta_n, b_n)_{n\in\mathbb{N}}$ (this can be done with only two states, using the method by Kari) and a transducer mapping each symbol (x, y) into $2y - x$, which can be done using only one state (this is just a relabelling).

There is indeed no reason that doing the transformation this way would work (in particular the equations given by Kari cannot be applied to this particular transducer and prove that there is indeed a tiling), and indeed it doesn't: we were able to prove that this particular Wang set does not, in fact, tile the plane.

Once this tileset was identified as belonging to the family of Kari tilesets, it is indeed easy to see that, should it tile the plane, it tiles a half plane starting from a word consisting only of 3.

We then started from a transducer $\mathcal{T}'$ that outputs a configuration with only the symbol 3, and build recursively $t_k = \mathcal{T}'\mathcal{T}^k$. It turns out that t_{31} (once reduced) is empty, which means that we cannot tile 31 consecutive rows starting from a word consisting only of 3.

Theorem 10.6. *There is no aperiodic Wang set with 10 tiles or less.*

Before removing unused transitions, t_{31} contains a path of 212 symbols 3. This means in particular that there exists a tiling of a rectangle of size 212×31 where the top and the bottom side are equal, thus a tiling of a infinite vertical strip of width 212 by this tiling, and thus a tiling of a square of size 212×212.

We want again to stress how much the simplification of the transducers by bisimulation was crucial. Our first proof that this tileset does not tile the plane did not use this and 3 months were needed to prove the result, generating sets of the order of 2^{32} (4 billion) tiles. Using bisimulation for the simplification of transducers, the result can be proven in 2 minutes, with the largest Wang set having 2^{26} (50 million) tiles.

The fact everything fall apart for $k = 31$ can be explained. If we identify $([0.5, 3]_{0.5\sim 3}, \times)$ with the unit circle $([0, 1]_{0\sim 1}, +)$, what f is doing is now just an addition (modulo 1) of $\frac{\log 2}{\log 2 + \log 3}$. Now $31 \frac{\log 2}{\log 2 + \log 3} = 11.992$ is near an integer, which means that $\mathcal{T}^{31}$ is "almost" the identity map. During the 30 first steps, our map $\mathcal{T}$ is able to deceive us and pretend it would tile the plane by using the degrees of freedom we have in the coding of the reals. For $k = 31$, this is not possible anymore.

It turns out that the exact same method can be used for the set of 12 tiles obtained starting from the set by Culik, and removing one tile. It corresponds to the same rotation, and we observe indeed the same behaviour: starting from a configuration of all 2, it is not possible to tile 31 consecutive rows:

Theorem 10.7. *The set of 13 tiles by Culik is minimal aperiodic: if any tile is removed from this set, it does not tile the plane anymore.*

Note that the situation is still not well understood and we can consider ourselves lucky to obtain the result: First, we have to execute the transducers in the good direction: $\mathcal{T}'\mathcal{T}^{-31}$ is nonempty. Furthermore, the next step when $\mathcal{T}^k$ returns near an integer is for $k = 106$, and no computer, using our technique, has enough memory to hope computing $\mathcal{T}^{106}$.

Conjecture 10.8. Every aperiodic tileset obtained by the method of Kari is minimal aperiodic.

10.4 An aperiodic Wang set of 11 tiles - Proof Sketch

Using the same method presented in the last section, we were able to enumerate and test sets of 11 tiles, and found a few potential candidates. Of these few candidates, two of them were extremely promising and we will indeed prove that they are aperiodic sets.

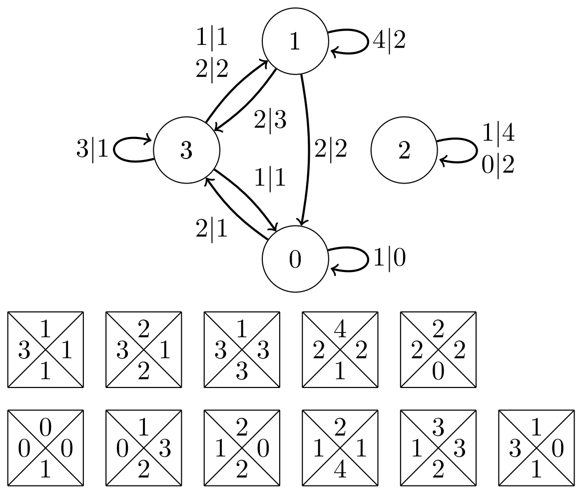


Figure 10.3 – Wang set $\mathcal{T}$.

These sets of tiles are presented in Figure 10.3 and 10.4. Both sets are very similar: the second one is obtained from the first one by collapsing the colors 4 and 0.

We focus now on the first set.

$\mathcal{T}$ is the union of two Wang sets, $\mathcal{T}_0$ and $\mathcal{T}_1$, of respectively 9 and 2 tiles. For $w \in \{0,1\}^* \setminus \{\epsilon\}$, let $\mathcal{T}_w = \mathcal{T}_{w[1]} \circ \mathcal{T}_{w[2]} \circ \ldots \mathcal{T}_{w[|w|]}$.

It can be seen by a easy computer check that every tiling by $\mathcal{T}$ can be decomposed into a tiling by transducers $\mathcal{T}_1\mathcal{T}_0\mathcal{T}_0\mathcal{T}_0\mathcal{T}_0$ and $\mathcal{T}_1\mathcal{T}_0\mathcal{T}_0\mathcal{T}_0$.

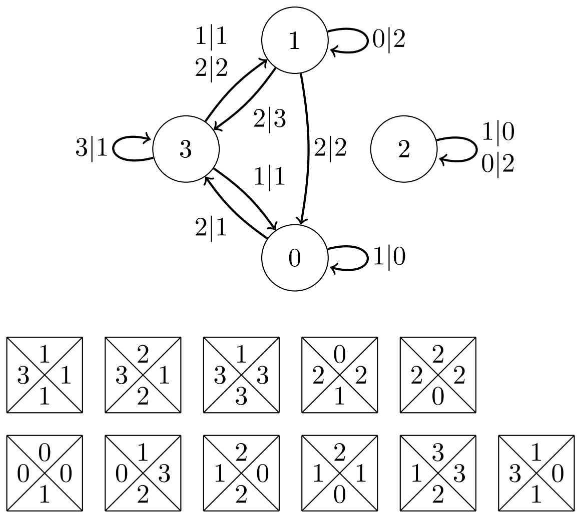


Figure 10.4 – Wang set $\mathcal{T}'$.

Simplifications of these two transducers, called $\mathcal{T}_a$ and $\mathcal{T}_b$ will be obtained in section 10.5.1 and are depicted in Fig. 10.5.

We then study the transducer $\mathcal{T}_D$ formed by the two transducers $\mathcal{T}_a$ and $\mathcal{T}_b$ and prove that there exists a tiling by $\mathcal{T}_D$, and that any tiling by $\mathcal{T}_D$ is aperiodic.

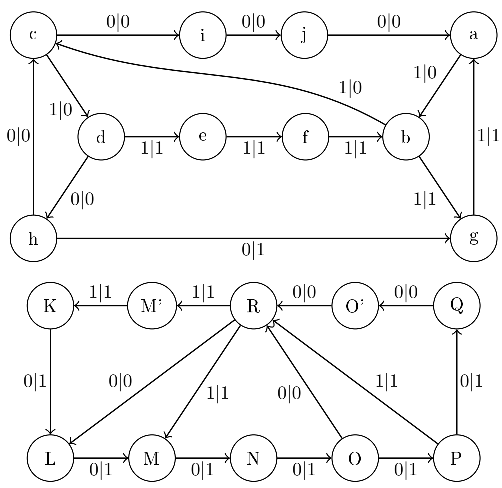


Figure 10.5 – $\mathcal{T}_D$, the union of $\mathcal{T}_a$ (top) and $\mathcal{T}_b$ (bottom).

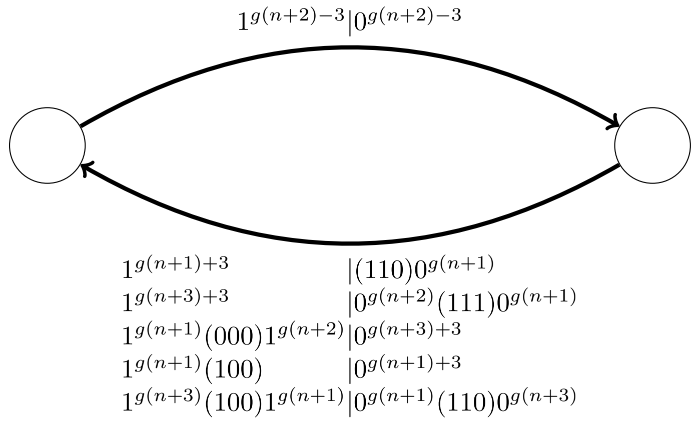
We will prove that the tileset is aperiodic by proving that any tiling is *substitutive*.

Let $u_{-2} = \epsilon, u_{-1} = a$, $u_0 = b, u_{n+2} = u_n u_{n-1} u_n$.

Let $g(n)$, $n \in \mathbb{N}$ be $(n+1)$-th Fibonacci number, that is $g(0) = 1$, $g(1) = 2$

and $g(n+2) = g(n) + g(n+1)$ for every $n \in \mathbb{N}$. Remark that u_n is of size $g(n)$.

T_n for n odd:



T_n for n even:

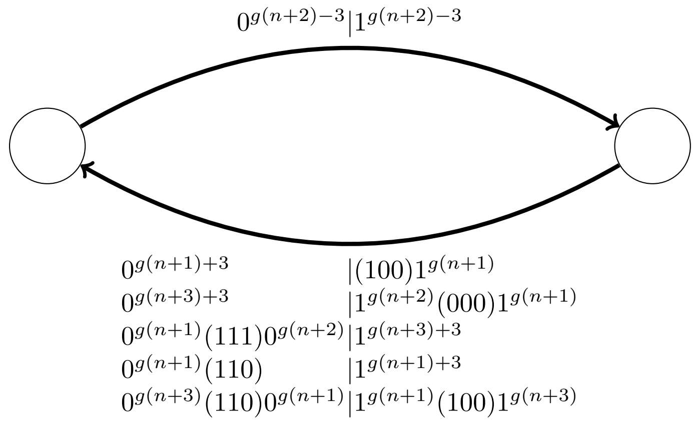


Figure 10.6 – The family of transducers T_n

Then we will prove that, for all n, any tiling by $\mathcal{T}_D$ is a tiling by $\mathcal{T}_{u_n}, \mathcal{T}_{u_{n+1}}, \mathcal{T}_{u_{n+2}}$. (This is obvious by definition for $n = -2, -1$). For this, we now introduce a family of transducers, presented in Fig 10.6, and we will prove

- We prove (section 10.5.2) that every tiling by $\mathcal{T}_D = \mathcal{T}_a \cup \mathcal{T}_b$ can be seen as a tiling by $\mathcal{T}_{u_0} \cup \mathcal{T}_{u_1} \cup \mathcal{T}_{u_2} = \mathcal{T}_b \cup \mathcal{T}_{aa} \cup \mathcal{T}_{bab}$.

- We prove (section 10.5.2) that $\mathcal{T}_{u_0}$, $\mathcal{T}_{u_1}$ and $\mathcal{T}_{u_2}$, when occuring in a tiling of the entire plane, can be simplified to obtain the three transducers T_0, T_1, T_2.

- We prove (section 10.6) that $T_{n+3} = T_{n+1} \circ T_n \circ T_{n+1}$ for all n, which proves that $\mathcal{T}_{u_n}$ can be simplified to obtain T_n

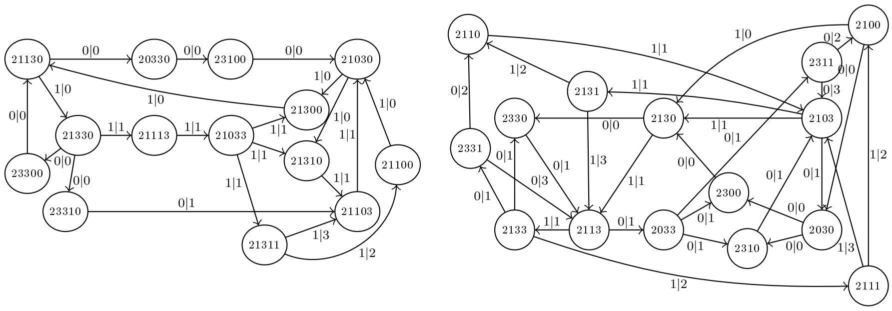


Figure 10.7 – $\mathcal{T}_A$, the union of $s(\mathcal{T}_{10000})$ (left) and $s(\mathcal{T}_{1000})$ (right).

- We then prove (section 10.7) that any tiling by $\mathcal{T}_{u_n}, \mathcal{T}_{u_{n+1}}$ and $\mathcal{T}_{u_{n+2}}$ can be rewritten as a tiling by $\mathcal{T}_{u_{n+1}}, \mathcal{T}_{u_{n+2}}, \mathcal{T}_{u_{n+3}}$, by replacing any block $\mathcal{T}_{u_{n+1}}\mathcal{T}_{u_n}\mathcal{T}_{u_{n+1}}$ by $\mathcal{T}_{u_{n+3}}$ (the difficulty is to prove that by doing this, there is no remaining occurence of $\mathcal{T}_{u_n}$).

- This proves in particular that every tiling is aperiodic.

- From the description of T_n, it is clear that the transducer T_n (hence $\mathcal{T}_{u_n}$) is nonempty. This implies that there exists a tiling of one row by $\mathcal{T}_{u_n}$, hence a tiling of $g(n)$ consecutive rows by $\mathcal{T}_a$ and $\mathcal{T}_b$, hence there exists a tiling of a plane.

Finally, we explain in section 10.7 how the same proof gives us also the aperiodicity of the set $\mathcal{T}'$.

10.5 From $\mathcal{T}$ to $\mathcal{T}_D$ then to T_0, T_1, T_2

10.5.1 From $\mathcal{T}$ to $\mathcal{T}_D$

Recall that our Wang set $\mathcal{T}$ can be seen as the union of two Wang sets, $\mathcal{T}_0$ and $\mathcal{T}_1$, of respectively 9 and 2 tiles.

For $w \in \{0,1\}^* \setminus \{\epsilon\}$, let $\mathcal{T}_w = \mathcal{T}_{w[1]} \circ \mathcal{T}_{w[2]} \circ \dots \mathcal{T}_{w[|w|]}$. The following facts can be easily checked by computer or by hand:

Fact 10.9. The transducers $s(\mathcal{T}_{11})$, $s(\mathcal{T}_{101})$, $s(\mathcal{T}_{1001})$ and $s(\mathcal{T}_{00000})$ are empty.

Thus, if t is a tiling by $\mathcal{T}$, then there exists a bi-infinite binary word $w \in \{1000, 10000\}^{\mathbb{Z}}$ such that $t(x,y) \in T(\mathcal{T}_{w[y]})$ for every $x, y \in \mathbb{Z}$. Let $\mathcal{T}_A = s(\mathcal{T}_{1000} \cup \mathcal{T}_{10000})$ (see Figure 10.7). There is a bijection between the tilings by $\mathcal{T}$ and the tilings by $\mathcal{T}_A$, and $\mathcal{T}$ is aperiodic if and only if $\mathcal{T}_A$ is aperiodic.

We see that the transducer $\mathcal{T}_A$ never reads 2, 3 nor 4. Thus the transitions that write 2, 3 or 4 are never used in a tiling by $\mathcal{T}$. Let $\mathcal{T}_B$ (see Figure 10.8) be the transducer $\mathcal{T}_A$ after removing these unused transitions, and deleting states that cannot appear in a tiling of a row (i.e. sources and sinks). Then t is a tiling by $\mathcal{T}_A$ if and only if t is a tiling by $\mathcal{T}_B$, and $\mathcal{T}_B$ is aperiodic if and only if $\mathcal{T}_A$ is.

Now we simplify a bit the transducer $\mathcal{T}_B$ using bisimulation. The states 23300 and 23310 have the same incoming transitions, hence can be coalesced

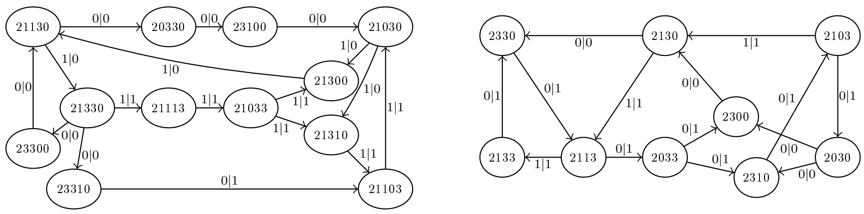


Figure 10.8 – $\mathcal{T}_B$ corresponds to $\mathcal{T}_A$ when unused transitions are deleted.

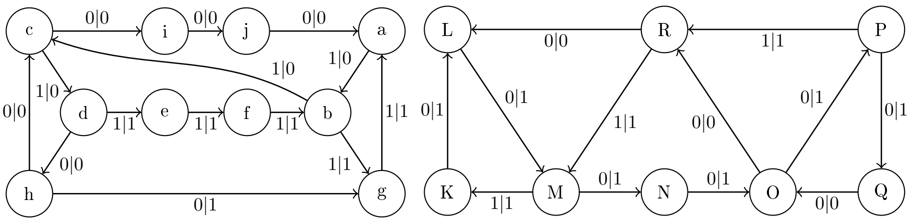


Figure 10.9 – $\mathcal{T}_C$ is the simplification of $\mathcal{T}_B$ by bisimulation.

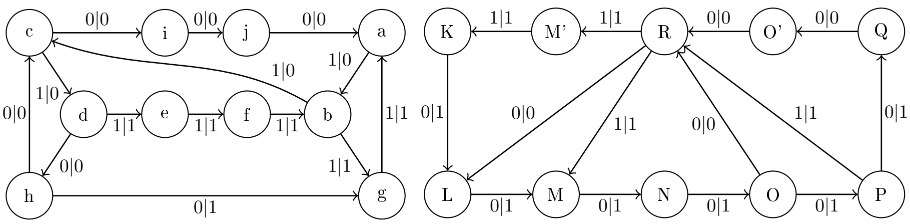


Figure 10.10 – $\mathcal{T}_D$ is the simplification of $\mathcal{T}_C$ using the fact that the successions of symbols 101 and 010 cannot appear. The transducers to the left and to the right are called respectively $\mathcal{T}_a$ and $\mathcal{T}_b$.

into one state The same goes for states 21300 and 21310, and for states 2300 and 2310. Once we coalesce all those states, we obtain the Wang set $\mathcal{T}_C$ depicted in Figure 10.9.

$\mathcal{T}_B$ and $\mathcal{T}_C$ are equivalent. Thus $\mathcal{T}_B$ is aperiodic if and only if $\mathcal{T}_C$ is aperiodic.

Proposition 10.10. *Let $(w_i)_{i\in\mathbb{Z}}$ be a bi-infinite sequence of bi-infinite binary words such that $w_i \mathcal{T}_C w_{i+1}$ for every $i \in \mathbb{Z}$. Then for every $i \in \mathbb{Z}$, w_i is $(010, 101)$-free.*

Proof. We consider the tiling in the other direction, and look at the transducer $(\mathcal{T}_C^{\mathrm{tr}})^3$. This transducer has 8 states (that corresponds respectively to 000, 001, ... 111) and a quick computer check shows that in this transducer the states 010 and 101 are respectively a source and a sink. As a consequence, these two states cannot appear in a tiling of the plane by $(\mathcal{T}_C^{\mathrm{tr}})^3$, hence 101 and 010 cannot appear in any line of a tiling by $\mathcal{T}_C$. ☐

In a tiling by $\mathcal{T}_C$, the transition from Q to O is never followed by a transition from O to P, otherwise it writes a 101. Similarly, a transition from M to K is never preceded by a transition from L to M, otherwise it reads a 010. Thus there is a bijection between tilings by $\mathcal{T}_C$ and tilings by $\mathcal{T}_D$ (Figure 10.10).

10.5.2 From $\mathcal{T}_D$ to T_0, T_1, T_2

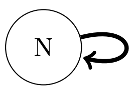
Let $\mathcal{T}_a$ and $\mathcal{T}_b$ be the two connected component of $\mathcal{T}_D$. For a word $w \in \{a, b\}^*$, let $\mathcal{T}_w = \mathcal{T}_{w[1]} \circ \mathcal{T}_{w[2]} \circ \ldots \mathcal{T}_{w[|w|]}$. The following fact can be easily checked by computer or by hand:

Fact 10.11. The transducers $s(\mathcal{T}_{bb})$, $s(\mathcal{T}_{aaa})$ and $s(\mathcal{T}_{babab})$ are empty.

It is a classical exercise to show that this implies that if t is a tiling by $\mathcal{T}_C$ then there exists a bi-infinite binary word $w \in \{b, aa, bab\}^{\mathbb{Z}}$ such that $t(x, y) \in T(\mathcal{T}_{w[y]})$ for every $y \in \mathbb{Z}$. That is, t is image of a tiling by $\mathcal{T}_b \cup \mathcal{T}_{aa} \cup \mathcal{T}_{bab}$.

We will now simplify the three transducers.

Case of $\mathcal{T}_b$. In $\mathcal{T}_b$, every path eventually go to the state "N". Thus $\mathcal{T}_b$ is equivalent to the following transducer (written in a compact form):



00000	10011
00000000	11100011
00111000	11111111
00110	11111
0000011000	1110011111
0010	1011
001000	111011
0000010	1110011
0011000	1011111

In the previous transducer, the last 4 transitions are never used in a tiling of the plane, since they read 010 or write 101. So we can simplify the transducer into:

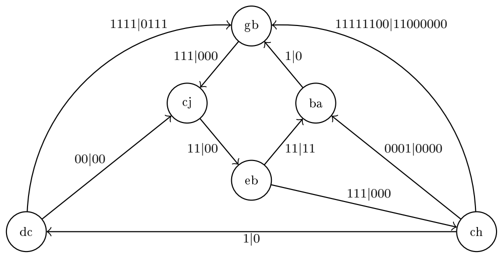


Figure 10.11 – $s(\mathcal{T}_{aa})$

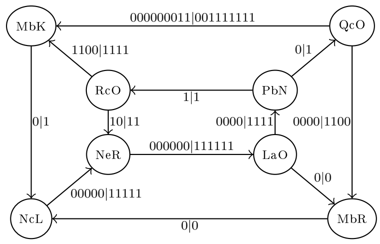
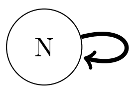
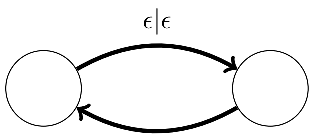


Figure 10.12 – $s(\mathcal{T}_{bab})$



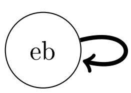
00000 |10011
00000000 |11100011
00111000 |11111111
00110 |11111
0000011000|1110011111

This transducer is equivalent to T_0, that we recall here for comparison:



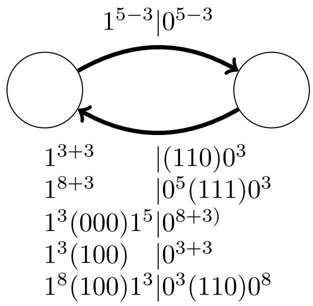
0^5 $|(100)1^2$
0^{5+3} $|1^3(000)1^2$
$0^2(111)0^3|1^{5+3}$
$0^2(110)$ $|1^{2+3}$
$0^5(110)0^2|1^2(100)1^5$

Case of $\mathcal{T}_{aa}$. The transducer $s(\mathcal{T}_{aa})$ is depicted in Figure 10.11 in a compact form. In this transducer, every path eventually go to the state "eb". Then $s(\mathcal{T}_{aa})$ is equivalent to the following transducer (written in a compact form):



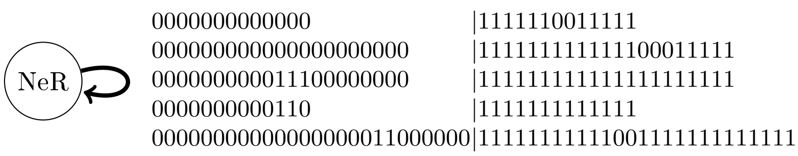
11111111 |11000000
1111111111111 |0000011100000
1110001111111 |0000000000000
11110011 |00000000
1111111110011111|0001100000000000

This transducer is clearly equivalent to T_1, that we recall for convenience:

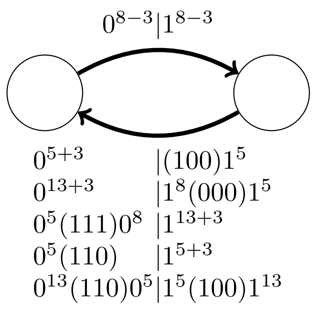


Case of $\mathcal{T}_{bab}$. The transducer $s(\mathcal{T}_{bab})$ is depicted in Figure 10.12.

In this transducer, every path eventually go to the state "NeR". Then $s(\mathcal{T}_{bab})$ is equivalent to the following transducer (wrote in a compact form):

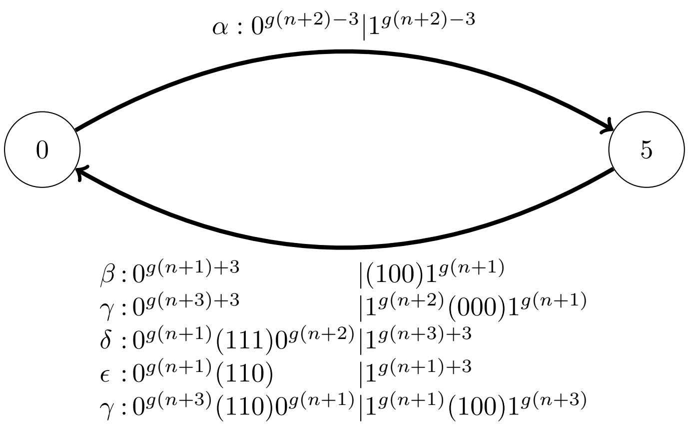


This transducer is clearly equivalent to T_2, that we recall for the reader convenience:

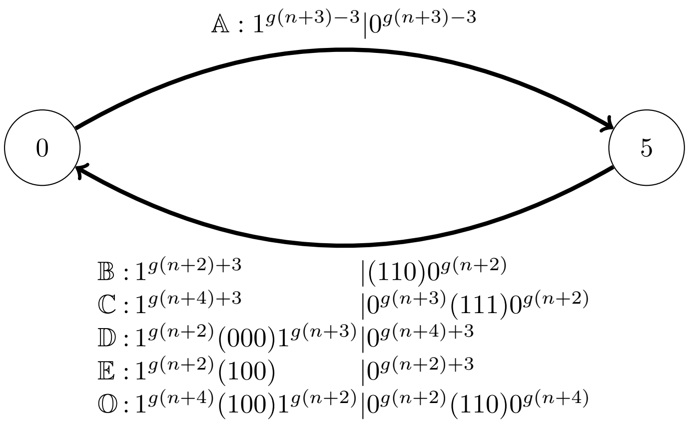


10.6 From T_n, T_{n+1}, T_{n+2} to $T_{n+1}, T_{n+2}, T_{n+3}$

For the reader convenience, we recall the definition of the family of transducers, and we introduce notations for the transitions T_n for n even:



T_{n+1} for n even:



Before going into the proof, we first give some remarks.

- T_n for n even and n odd are essentially similar. This means it is sufficient to prove that $T_{n+3} = T_{n+1} \circ T_n \circ T_{n+1}$ for n even, and the result for n odd follows.

- Apply the following transformation to T_n: Change input and output, and reverse the edges: reverse the direction and mirror (reverse) the words, and exchange the symbols 0 and 1. Then we obtain T_n again (for n even, with β playing the role of ϵ, δ the role of γ, and α and ω their own role). This internal symmetry will be used a lot in the proofs.

- All transitions are symmetric and easy to understand, except the self-symmetric tiles ω and $\mathbb{O}$. These transitions actually cannot occur in the tiling of the plane, but a transition of shape ω or $\mathbb{O}$ large enough can appear in a finite strip large enough. It means it is not possible to do the proof without speaking about these transitions, even if they cannot appear in a tiling of the plane.

We now proceed to prove the result. As said before, we now suppose that n is even, and we will look at the sequence of transducers $T_{n+1} \circ T_n \circ T_{n+1}$.

The following table represent the possible distance between two consecutive markers (i.e. 000 and 100) as inputs of T_{n+1}.

First Marker	Second Marker	Distance	
(000) from D	(000) from D	$g(n{+}5)$	
(000) from D	(100) from E	$g(n{+}5)$	
(000) from D	(100) from O	$g(n{+}5){+}g(n{+}3)$	
(100) from E	(000) from D	$g(n{+}4)$	
(100) from E	(100) from E	$g(n{+}4)$	$+ag(n{+}4){+}bg(n{+}5)$
(100) from E	(100) from O	$g(n{+}5)$	$a, b \in \mathbb{N}$
(100) from O	(000) from D	$g(n{+}4){+}g(n{+}2)$	
(100) from O	(100) from E	$g(n{+}4){+}g(n{+}2)$	
(100) from O	(100) from O	$2g(n{+}4)$	

To prove the main result, we will prove that the transitions in the transducer T_n (when surrounded by transducers T_{n+1}) must be done in a certain order.

In the following, we deliberately omit the transition α: When we say that $\gamma\beta$ cannot appear, we mean that it is impossible to see successively the transitions γ, then α, then β in a run of the transducer T_n (when surrounded by transducers T_{n+1}).

Lemma 10.12. *The following words cannot appear:*

- $\gamma\omega, \gamma\gamma, \gamma\beta, \beta\omega, \beta\beta, \beta\epsilon\beta, \gamma\epsilon\beta,\ \beta\delta\epsilon\beta,\ \gamma\delta\epsilon\beta$
- $\omega\delta, \delta\delta, \epsilon\delta, \omega\epsilon, \epsilon\epsilon, \epsilon\beta\epsilon, \epsilon\beta\delta,\ \epsilon\beta\gamma\epsilon,\ \epsilon\beta\gamma\delta$

Proof. All the following successions of transitions are impossible due to the input constraints on T_{n+1} :

Case	Why it is impossible
$\gamma\omega$	(000) and (100) separated by $g(n+1)+g(n+3)$
$\gamma\gamma$	(000) and (000) separated by $g(n+4)$
$\gamma\beta$	(000) and (100) separated by $g(n+3)$
$\beta\omega$	(100) and (100) separated by $g(n+1)+g(n+3)$
$\beta\beta$	(100) and (100) separated by $g(n+3)$
$\beta\epsilon\beta$	(100) and (100) separated by $2g(n+3)$
$\gamma\epsilon\beta$	(000) and (100) separated by $2g(n+3)$
$\beta\delta\epsilon\beta$	(100) and (100) separated by $2g(n+4)+g(n+1)$
$\gamma\delta\epsilon\beta$	(000) and (100) separated by $2g(n+4)+g(n+1)$

All others cases follow by symmetry. □

Lemma 10.13. *ω cannot appear.*

Proof. Case disjunction on what appears before:

Case	Why it is impossible
$\beta\omega$	see above
$\gamma\omega$	see above
$\beta\delta\omega$	(100) and (100) separated by $g(n+4)+g(n+3)+g(n+1)$
$\gamma\delta\omega$	(000) and (100) separated by $g(n+4)+g(n+3)+g(n+1)$
$\beta\epsilon\omega$	(100) and (100) separated by $g(n+4)+2g(n+1)$
$\gamma\epsilon\omega$	(000) and (100) separated by $g(n+4)+2g(n+1)$
$\beta\delta\epsilon\omega$	(100), (100) separated by $g(n+5)+g(n+3)+g(n+1)=2g(n+4)+2g(n+1)$
$\gamma\delta\epsilon\omega$	(000), (100) separated by $g(n+5)+g(n+3)+g(n+1)=2g(n+4)+2g(n+1)$

□

Lemma 10.14. *$\mathbb{O}$ cannot appear.*

Proof. Suppose that $\mathbb{O}$ appear in the top transducer (i.e. the transducers with input T_n). This means the (100) marker is generated, the only possibility being by β.

We prove there is no possibility to find transitions after this β.

Case	Why it is impossible starting from $\mathbb{O}$
$\beta\gamma$	(100) and (000) separated by $g(n+4)$
$\beta\delta\beta$	(100) and (100) separated by $g(n+4)+g(n+3)$
$\beta\delta\gamma$	(100) and (000) separated by $g(n+4)+g(n+1)+g(n+3)$
$\beta\delta\epsilon\beta$	(100) and (100) separated by $2g(n+4)+g(n+1)$
$\beta\delta\epsilon\gamma$	(100) and (000) separated by $2g(n+4)+g(n+3)$
$\beta\epsilon\gamma$	(100) and (000) separated by $g(n+5)$

By symmetry, $\mathbb{O}$ cannot appear in the bottom transducer. □

Now that $\mathbb{O}$ has disappeared, the possible distances between the markers are greatly simplified

First Marker	Second Marker	Distance	
(000)	(000)	$g(n+5)$	$+ag(n+4)+bg(n+5)$, $a,b\in\mathbb{N}$
(000)	(100)	$g(n+5)$	
(100)	(000)	$g(n+4)$	
(100)	(100)	$g(n+4)$	

Lemma 10.15. *The following words do not appear: $\beta\epsilon$, $\epsilon\beta$ $\beta\delta\beta$, $\delta\gamma\delta$, as well as $\epsilon\gamma\epsilon$ and $\gamma\delta\gamma$*

Proof. $\beta\epsilon$ should be followed by γ which leads to (100) and (000) separated by $g(n+5)$.

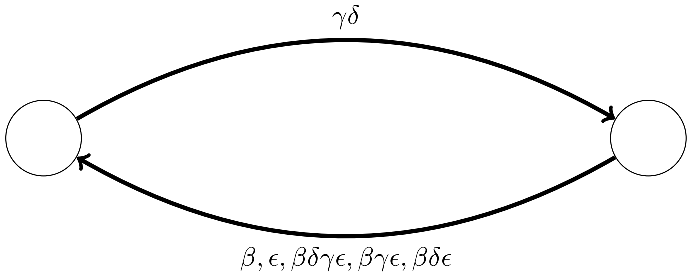
$\epsilon\beta$ should be preceded by a δ, which cannot be preceded by anything.

Case	Why it is impossible
$\beta\delta\beta$	(100), (100) separated by $g(n+4)+g(n+3)$
$\gamma\delta\gamma$	(000), (000) separated by $g(n+5)+g(n+2)$

The last two follow by symmetry.

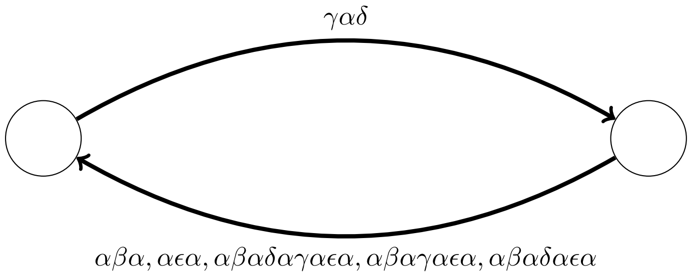
□

Lemma 10.16. *Every infinite path on the transducer T_n can be written as paths on the following graph:*



Proof. Clear: all other words are forbidden by the previous lemmas □

Recall that in this picture, words α have been forgotten. We now rewrite it adding the transitions α.

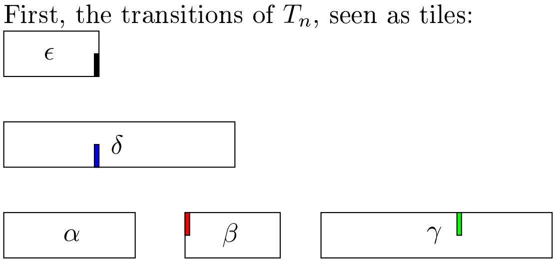


All transitions in the picture will be called *meta-transitions*.

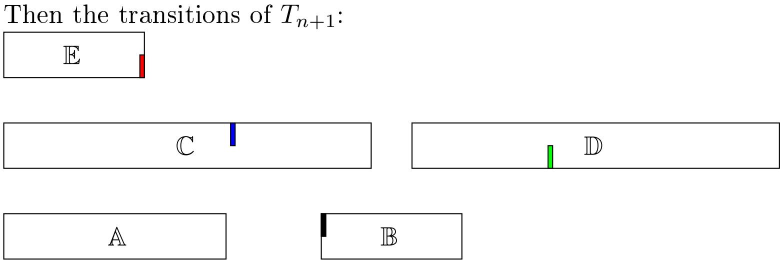
We now have a more accurate description of the behaviour of the transducer T_n when surrounded by transducers T_{n+1}. This will be sufficient to prove the results. We will see indeed that each of the six meta-transitions depicted can be completed in only one way by transitions of T_{n+1}. This will give us six tiles, which (almost) correspond to the transitions of T_{n+3}.

We will use drawings to prove the result. Let first draw all tiles: The pictures will be self-explanatory.

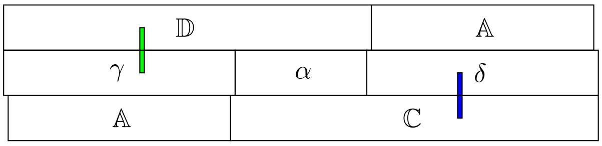
First, the transitions of T_n, seen as tiles:



Then the transitions of T_{n+1}:

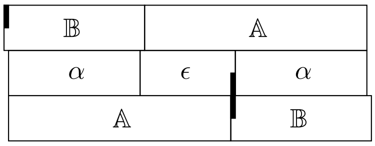


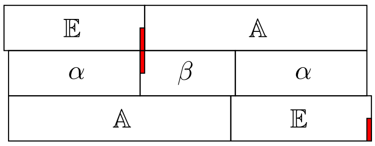
We now first look at $\gamma\delta$. By necessity, the following transitions of T_{n+1} should surround it:



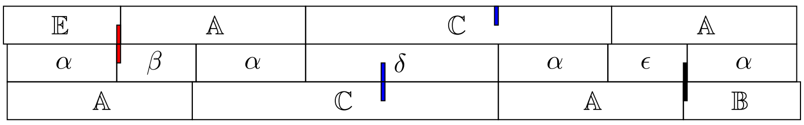
Note that the three transducers are aligned (up to a a shift of ± 3) when $\gamma\alpha\delta$ is present. As all other meta-transitions are enclosed by the meta-transition $\gamma\alpha\delta$, This means that in an execution of $T_{n+1} \circ T_n \circ T_{n+1}$, every other meta-transition should be surrounded above and below by transitions of T_{n+1} that almost align with it. Moreover the transitions of T_{n+1} below should begin by $\mathbb{A}$ and the transitions of T_{n+1} above should end with $\mathbb{A}$. It turns out that there is only one way to do this for any of the meta-transitions.

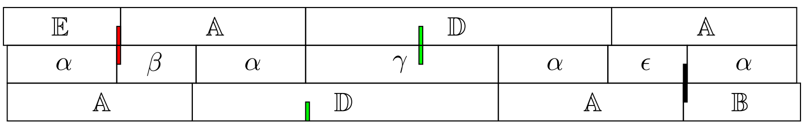
This gives for ϵ and β:



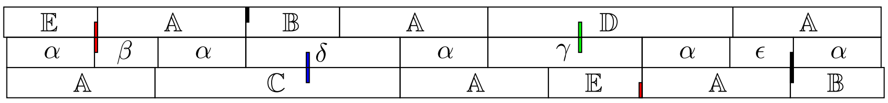


This gives for $\beta\gamma\epsilon$ and $\beta\delta\epsilon$:

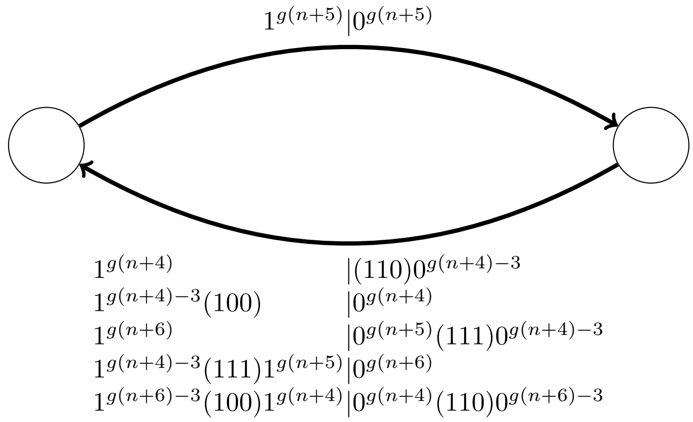




And the piece de resistance $\beta\delta\gamma\epsilon$:



We now look at the transducer T' we obtain with the preceding six pieces. Remark that $T' = T_n \circ T_{n+1} \circ T_n \circ \sigma^3$ where σ is the shift:



It is easy to see that T' is exactly T_{n+3} up to a shift of 3.

Theorem 10.17. $T_{n+3} = T_n \circ T_{n+1} \circ T_n$.

10.7 End of the proof

10.7.1 Aperiodicity of $\mathcal{T}$

Theorem 10.18. *Every infinite composition of the tranducers T_n, T_{n+1}, T_{n+2} can be rewritten as a composition of transducers $T_{n+1}, T_{n+2}, T_{n+3}$ by replacing every block $T_{n+1} \circ T_n \circ T_{n+1}$ by T_{n+3}. In particular, every tiling by T_0, T_1 and T_2 is aperiodic.*

Proof. It is easy to see, given the inputs of T_n, T_{n+1} and T_{n+2}, that every T_n should be bordered by the transducers T_{n+1}.

It therefore remains to show that $T_{n+1} \circ T_n \circ T_{n+1} \circ T_n \circ T_{n+1}$ cannot appear.

By the previous section, T_n, when bordered by T_{n+1} on both sides, can be rewritten as concatenations of blocks of the following five types: $\beta\gamma\delta$, $\epsilon\gamma\delta$, $\beta\delta\gamma\epsilon\gamma\delta$, $\beta\gamma\epsilon\gamma\delta$ and $\beta\delta\epsilon\gamma\delta$.

However, as $T_{n+1} \circ T_n \circ T_{n+1} \circ T_n \circ T_{n+1} = T_{n+3} \circ T_n \circ T_{n+1}$, the block $\epsilon\gamma\delta$ (and any block containing it) cannot appear in the execution of the transducer $T_n$is impossible, as T_{n+3} does not produce any input that where 100 and 000 are that close. So the only block that remain possibly is $\beta\gamma\delta$. But T_{n+3} does not produce any input where 000 and 000 are at distance $g(n+6)$. □

Corollary 10.19. *The Wang set corresponding to the transducer $T_0 \cup T_1 \cup T_2$ is aperiodic*

Corollary 10.20. *The Wang set $\mathcal{T}_D = \mathcal{T}_a \cup \mathcal{T}_b$ is aperiodic. Furthermore, the set of words $u \in \{a, b\}^\star$ s.t. the sequence of transducers $\mathcal{T}_u$ appear in a tiling of the plane is exactly the set of factors of the Fibonacci word (*i.e. *the fixed point of the morphism $a \to ab, b \to a$), i.e. the set of factors of sturmian words of slope $1/\phi$, for ϕ the golden mean.*

The set of biinfinite words $u \in \{a, b\}^\mathbb{Z}$ s.t $\mathcal{T}_u$ represents a valid tiling of the plane are exactly the sturmian words of slope $1/\phi$.

See [43] for some references on sturmian words.

Proof. The sequence of words u_n we defined is the sequence of singular factors of the Fibonacci word (see for example [148]). Thus, on tilings by $\mathcal{T}_a \cup \mathcal{T}_b$, the vertical sequence on $\{a, b\}$ have the same set of factors that the Fibonacci word. $\square$

Corollary 10.21. *The Wang set $\mathcal{T}$ is aperiodic. Furthermore, the set of words $u \in \{0,1\}^\star$ s.t. the sequence of transducers $\mathcal{T}_u$ appear in a tiling of the plane is exactly the set of factors of sturmian words of slope $\phi/(5\phi - 1)$, for ϕ the golden mean.*

The set of biinfinite words $u \in \{0,1\}^\mathbb{Z}$ s.t $\mathcal{T}_u$ represents a valid tiling of the plane are exactly the sturmian words of slope $\phi/(5\phi - 1)$.

Proof. Let ψ be the morphism $a \mapsto 10000, b \mapsto 1000$. The set of all words $u \in \{0,1\}^\mathbb{Z}$ that can appear in a tiling of the whole plane are exactly the image by ψ of the sturmian words over the alphabet $\{a, b\}$ of slope $1/\phi$.

It is well known that the image of a sturmian word by ψ is again a sturmian word, see [43, Corollary 2.2.19], where $\psi = \tilde{G}^3 D$ (with $\{a, b\}$ instead of $\{0, 1\}$ as input alphabet). The derivation of the slope is routine. $\square$

10.7.2 Aperiodicity of $\mathcal{T}'$

Recall that $\mathcal{T}'$ is the Wang set from Figure 10.4. This Wang set is obtained from $\mathcal{T}$, by merging two vertical colors: 0 and 4 in $\mathcal{T}$ become 0 in $\mathcal{T}'$. Thus every tiling of $\mathcal{T}$ can be turned into a tiling of $\mathcal{T}'$, and $\mathcal{T}'$ tiles the plane. We will show in the sequel that every tiling of $\mathcal{T}'$ can be turned into a tiling of $\mathcal{T}$, and thus every tiling of $\mathcal{T}'$ is aperiodic.

$\mathcal{T}'$ is the union of two Wang sets $\mathcal{T}'_0$ and $\mathcal{T}'_1$ of respectively 9 and 2 tiles. The following facts can be easily checked by computer. For $w \in \{0,1\}^* \setminus \{\epsilon\}$, let $\mathcal{T}'_w = \mathcal{T}'_{w[1]} \circ \mathcal{T}'_{w[2]} \circ \dots \mathcal{T}'_{w[|w|]}$.

Fact 10.22. The transducers $s(\mathcal{T}'_{111})$, $s(\mathcal{T}'_{101})$, $s(\mathcal{T}'_{1001})$, $s(\mathcal{T}'_{1000001})$, $s(\mathcal{T}'_{10000001})$, $s(\mathcal{T}'_{100000001})$, $s(\mathcal{T}'_{000000000})$, $s(\mathcal{T}'_{000011})$, $s(\mathcal{T}'_{110000})$ and $s(\mathcal{T}'_{1100011})$ are empty.

Thus, if t is a tiling by $\mathcal{T}'$ then there exists a bi-infinite binary word $w \in \{1000, 10000, 100011000, 100000000\}^\mathbb{Z}$ such that $t(x, y) \in T(\mathcal{T}'_{w[y]})$ for every $x, y \in \mathbb{Z}$.

Let $\mathcal{T}'_A = s(\mathcal{T}'_{1000} \cup \mathcal{T}'_{10000} \cup \mathcal{T}'_{100000000} \cup \mathcal{T}'_{100011000})$. As before, $\mathcal{T}'_A$ has unused transitions (those which writes 2 or 3). Once deleted, with states that cannot appear in a tiling of a row, we obtain $\mathcal{T}'_B$. $\mathcal{T}'_B$ has 4 connected components: two were already present in $\mathcal{T}$: $\mathcal{T}_a$ and $\mathcal{T}_b$, the third one $\mathcal{T}_c$ is a subset of $\mathcal{T}'_{100000000}$, and the last one $\mathcal{T}_d$ is a subset of $\mathcal{T}'_{100011000}$.

Proposition 10.23. *$\mathcal{T}'_{11}$ is isomorphic to a subset of $\mathcal{T}'_{01}$, and $\mathcal{T}'_{100000}$ is isomorphic to a subset of $\mathcal{T}'_{100001}$.*

Proof. $\mathcal{T}'_{11}$ is the transducer with one state, which reads 1 and writes 2. $\mathcal{T}'_{01}$ has also a loop that reads 1 and writes 2: the transition $(02, 02, 1, 2)$. $\mathcal{T}'_{100000}$ and $\mathcal{T}'_{100001}$ are depicted in Figure 10.13 (in a compact form). $\mathcal{T}'_{100000}$ is isomorphic to the subset of $\mathcal{T}'_{100001}$ drawn in bold. $\square$

Corollary 10.24. *$\mathcal{T}_c$ and $\mathcal{T}_d$ are both isomorphic to a subset of $\mathcal{T}_a \circ \mathcal{T}_b$.*

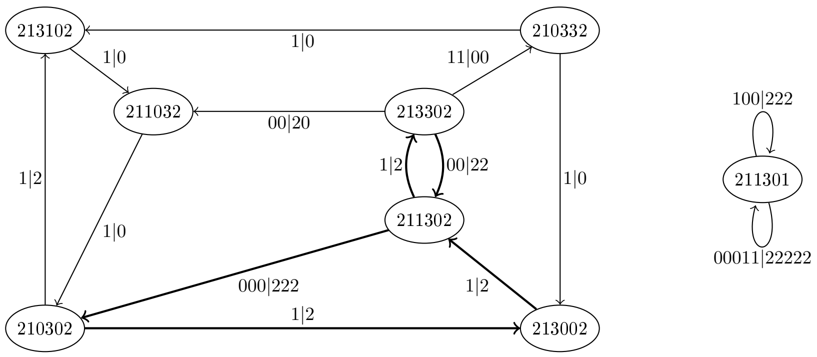


Figure 10.13 – $\mathcal{T}'_{100001}$ (left) and $\mathcal{T}'_{100000}$ (right).

A tiling of $\mathcal{T}'_B$ can thus be turned into a tiling of $\mathcal{T}_B$, by substituting every tile from $\mathcal{T}_c$ (resp. $\mathcal{T}_d$) by two tiles, one from $\mathcal{T}_a$ and one from $\mathcal{T}_b$.

Theorem 10.25. *The Wang set $\mathcal{T}'$ is aperiodic.*

Proof. The Wang set $\mathcal{T}'$ is aperiodic if and only if $\mathcal{T}'_B$ is aperiodic. Suppose that $\mathcal{T}'_B$ is not aperiodic. We know that $\mathcal{T}'$, and thus $\mathcal{T}'_B$ tile the plane. Take a periodic tiling by $\mathcal{T}'_B$. This tiling can be turned into a tiling of $\mathcal{T}_B$ by the Corollary 10.24. Thus $\mathcal{T}_B$ has a periodic tiling, contradiction. ☐

10.7.3 Concluding remarks

- The reader may regret that our substitutive system starts from $\mathcal{T}_b \cup \mathcal{T}_{aa} \cup \mathcal{T}_{bab}$ and not from $\mathcal{T}_a \cup \mathcal{T}_b \cup \mathcal{T}_{aa}$, or even from $\mathcal{T}_a \cup \mathcal{T}_b$. We do not know if this is possible. Our definition of T_n certainly does not work for $n = -1$, and the natural generalization of it is not equivalent to $\mathcal{T}_a$. This is somewhat obvious, as T_n (for $n \geq 0$) cannot be composed with itself, whereas $\mathcal{T}_a$ should be composed with itself to obtain $\mathcal{T}_{aa}$.

- $\mathcal{T}_a$ and $\mathcal{T}_b$ both have the properties that they are time symmetric: If we reverse the directions of all edges, exchange inputs and outputs, and exchange 0 and 1, we obtain an equivalent transducer (it is obvious for $\mathcal{T}_b$ and become obvious for $\mathcal{T}_a$ if we write it in a compact form without the states h and g). This property was used to simplify the proof that the sequence (T_n) is a recursive sequence, but we do not know whether it can be used to simplify the whole proof.

- While we gave a sequence of transducers T_n, it is of course possible to give another sequence of transducers, say U_n, which are equivalent to T_n, and thus with the same properties. Our sequence T_n has nice properties, in particular the symmetry explained above and its short number of transitions, but has the drawback that the substitution once seen geometrically has small bumps due to the fact that the tiles are aligned only up to ± 3. It is possible to find a sequence U_n for which this does not appear, by splitting some transitions of T_n into transitions of size $g(k)$ and transitions of size exactly 3. However this makes the proof that the sequence is recursive harder. We think our sequence T_n reaches a nice compromise.

- We do now know if it is possible to obtain the result directly on the original tileset $\mathcal{T}$ rather than $\mathcal{T}_D$. A difficulty is that $\mathcal{T}$ is not purely substitutive (due for example to the fact that no sturmian word of slope $\phi/(5\phi-1)$ is purely morphic): What we could obtain at best is that tilings by $\mathcal{T}$ are images by some map ϕ of some substitutive tilings (which is more or less what we obtain in our proof).

- We have now obtained a large number of Wang sets with 11 tiles which are candidates for aperiodicity. The reader might ask why we choose to investigate this particular one. The reason is that, for this particular tileset $\mathcal{T}$, it is very easy for a computer to produce the transducer for $\mathcal{T}^k$ even for large values of k ($k = 1000$). For comparison, for almost all other tilesets, we were not able to reach even $k = 30$. This suggested this tileset had some particular structure. We will not give here a full bestiary of all our candidates, but we will say that a large number of them are tileset corresponding to the method of Kari, with one tile or more omitted. With the method we described previously we were able to prove that some of them do not tile the plane, but the method did not work on all of them. We have found for now only three tilesets which were likely to be substitutive or nearly substitutive, of which two are presented in this article.

- Experimental results tend to support the following conjecture

  Conjecture 10.26. Let $f(n)$ be the smallest k s.t. every Wang set of size n that does not tile the plane does not tile a square of size k. Let $g(n)$ be the smallest k s.t. every Wang set of size n that tiles the plane periodically does so with a period $p \leq k$.

  Then $g(n) \leq f(n)$ for all n.

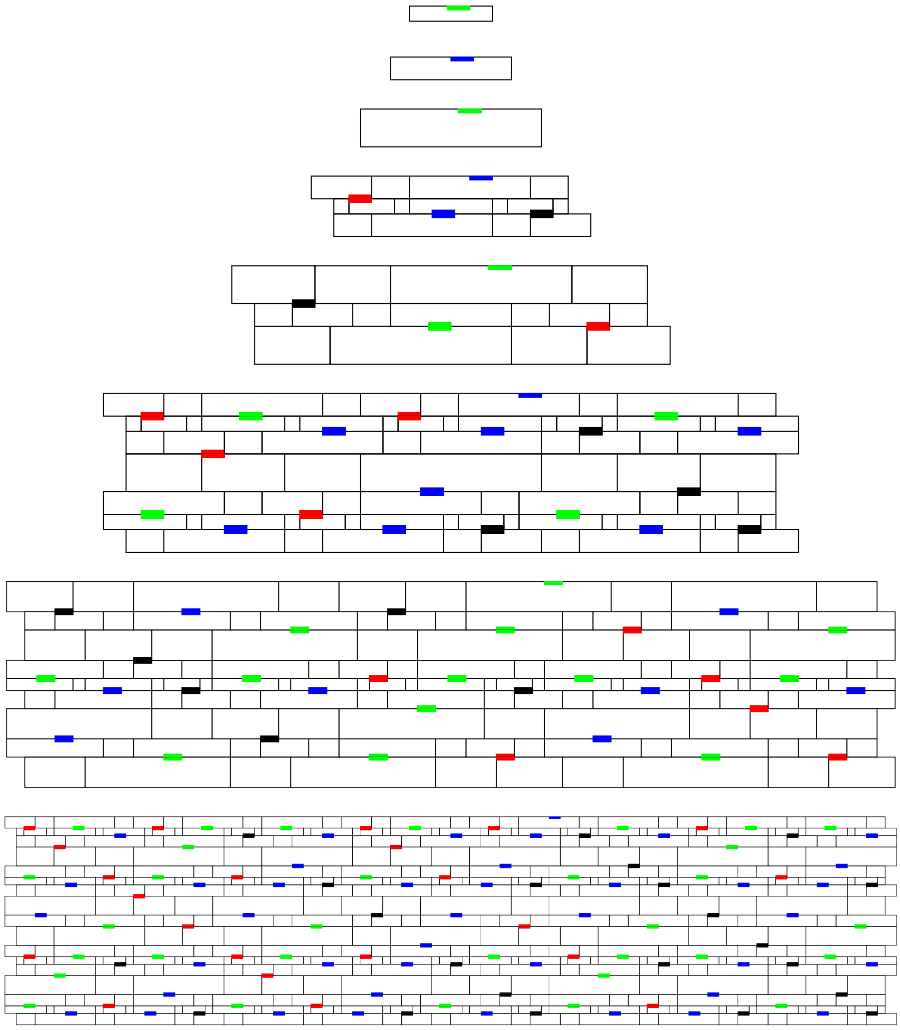


Figure 10.14 – Representation of the meta-tile γ (resp. C if n is odd) of T_n as tiles of $T_0 \uplus T_1 \uplus T_2$ for $n = 0, 1, 2, 3, 4, 5, 6, 7$.

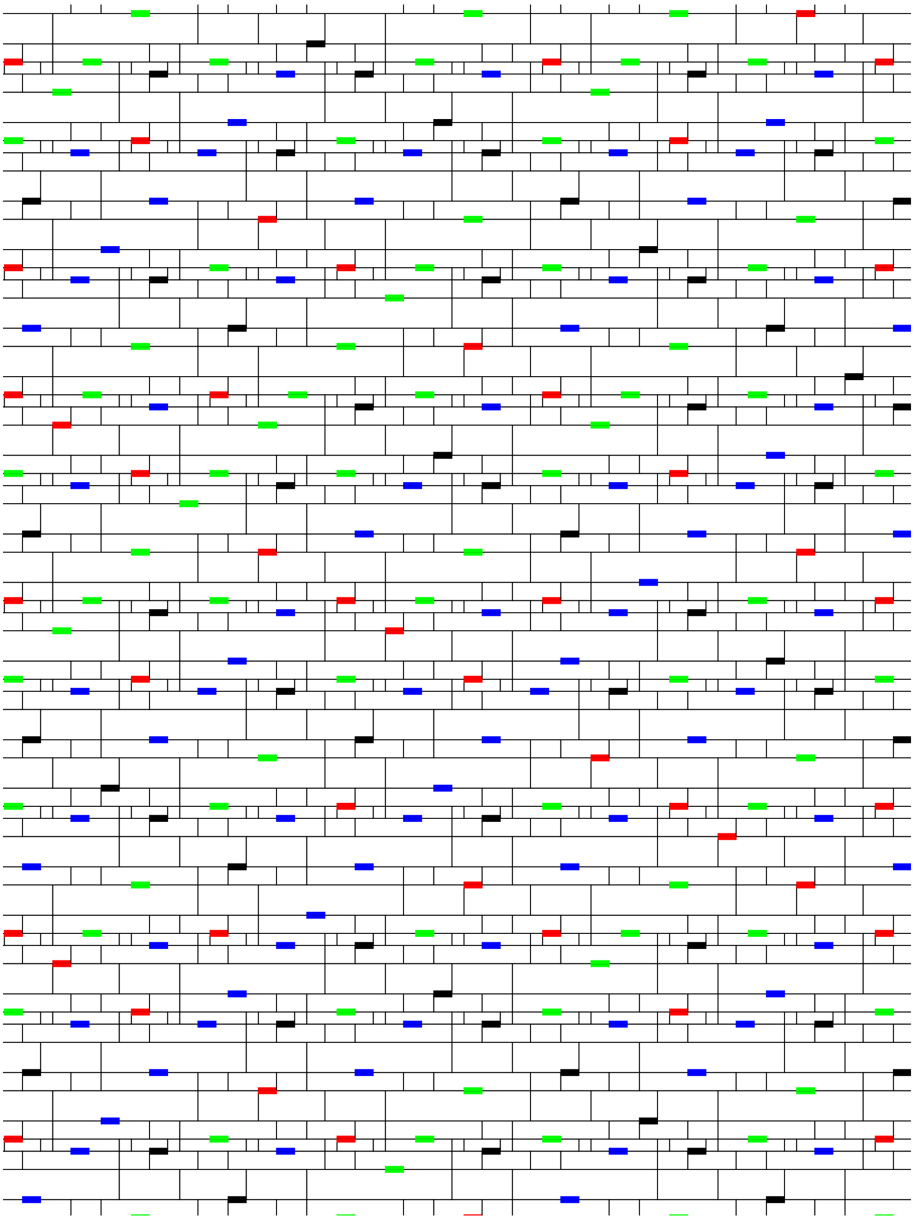


Figure 10.15 – A fragment of a tiling by the transducers T_0, T_1, T_2.

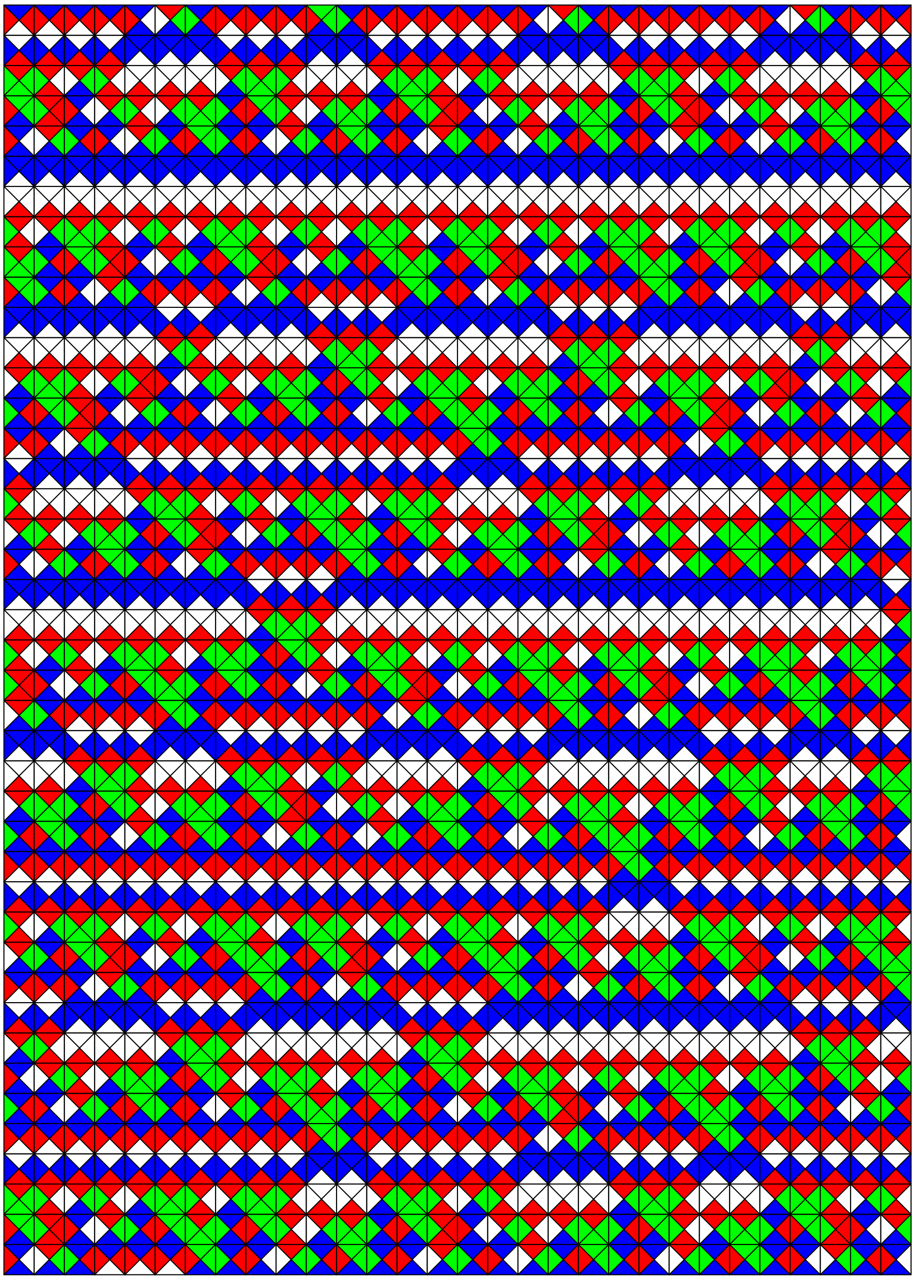


Figure 10.16 – A fragment of a tiling by $\mathcal{T}'$, with (0,1,2,3)=(white,red,blue,green).

Publications

[1] P. Allen, V. Lozin and M. Rao. "Clique-Width and the Speed of Hereditary Properties". In: *The Electronic Journal of Combinatorics* 16(1) (2009), #R35.

[2] G. Badkobeh, M. Crochemore and M. Rao. "Finite repetition threshold for large alphabets". In: *RAIRO-Theor. Inf. Appl.* 48.4 (2014), pp. 419–430.

[3] N. Bédaride and M. Rao. "Regular simplices and periodic billiard orbits". In: *Proc. Amer. Math. Soc.* 142 (2014), pp. 3511–3519.

[4] O. Bodini, T. Fernique, M. Rao and É. Rémila. "Distances on rhombus tilings". In: *Theoretical Computer Science* 412.36 (2011). Advances in Discrete Geometry for Computer Imagery, pp. 4787–4794.

[5] H. L. Bodlaender, A. Brandstädt, D. Kratsch, M. Rao and J. Spinrad. "On algorithms for (P_5,gem)-free graphs". In: *Theoretical Computer Science* 349.1 (2005), pp. 2–21.

[6] B.-M. Bui-Xuan, M. Habib and M. Rao. "Tree-representation of set families and applications to combinatorial decompositions". In: *European Journal of Combinatorics* 33.5 (2012). EuroComb '09, pp. 688–711.

[7] P. Charbit, M. Habib, V. Limouzy, F. de Montgolfier, M. Raffinot and M. Rao. "A note on computing set overlap classes". In: *Information Processing Letters* 108.4 (2008), pp. 186–191.

[8] J. Daligault, D. Gonçalves and M. Rao. "Diamond-free circle graphs are Helly circle". In: *Discrete Mathematics* 310.4 (2010), pp. 845–849.

[9] J. Daligault, M. Rao and S. Thomassé. "Well-Quasi-Order of Relabel Functions". In: *Order* 27.3 (2010), pp. 301–315.

[10] D. Gonçalves, A. Pinlou, M. Rao and S. Thomassé. "The Domination Number of Grids". In: *SIAM Journal on Discrete Mathematics* 25.3 (2011), pp. 1443–1453.

[11] E. Jeandel and M. Rao. "An aperiodic set of 11 Wang tiles". In: *Manuscript* (2015).

[12] M. M. Kanté and M. Rao. "The Rank-Width of Edge-Coloured Graphs". In: *Theory of Computing Systems* 52.4 (2013), pp. 599–644.

[13] M. M. Kanté and M. Rao. "Directed Rank-Width and Displit Decomposition". In: *WG 2009, LNCS*. Vol. 5911. 2010, pp. 214–225.

[14] J. Karhumäki, S. Puzynina, M. Rao and M. A. Whiteland. "On cardinalities of k-abelian equivalence classes". In: *Theoretical Computer Science* 658, Part A (2017), pp. 190–204.

[15] R. Kolpakov and M. Rao. "On the number of Dejean words over alphabets of 5, 6, 7, 8, 9 and 10 letters." In: *Theoretical Computer Science* 412.46 (2011), pp. 6507–6516.

[16] G. Kucherov, P. Ochem and M. Rao. "How many square occurrences must a binary sequence contain?" In: *The Electronic Journal of Combinatorics* 10.1 (2003), R12.

[17] F. de Montgolfier and M. Rao. "The bi-join decomposition". In: *Electronic Notes in Discrete Mathematics* 22 (2005), pp. 173–177.

[18] M. Müller, S. Puzynina and M. Rao. "On Shuffling of Infinite Square-Free Words". In: *The Electronic Journal of Combinatorics* 22(1) (2015).

[19] P. Ochem and M. Rao. "Minimum frequencies of occurrences of squares and letters in infinite words". In: *Mons Days of Theoretical Computer Science. Mons, August 27-30.* 2008.

[20] P. Ochem and M. Rao. "Another remark on the radical of an odd perfect number". In: *Fibonacci Quart* 52(3) (2014), pp. 215–217.

[21] P. Ochem and M. Rao. "Odd perfect numbers are greater than 10^{1500}". In: *Mathematics of Computation* 81 (2012), pp. 1869–1877.

[22] P. Ochem and M. Rao. "On the number of prime factors of an odd perfect number". In: *Mathematics of Computation* 83 (2014), pp. 2435–2439.

[23] M. Rao. "On some generalizations of abelian power avoidability". In: *Theoretical Computer Science* 601 (2015), pp. 39–46.

[24] M. Rao and M. Rosenfeld. "Avoidability of long k-abelian repetitions". In: *Mathematics of Computation* 85 (2016), pp. 3051–3060.

[25] M. Rao and M. Rosenfeld. "Avoiding two consecutive blocks of same size and same sum over $\mathbb{Z}^2$". In: *Manuscript* (2016).

[26] M. Rao. "Trucs et bidules sur la séquence de Kolakoski". In: *http://www.arthy.org/kola/kola.php* (2012).

[27] M. Rao. "Clique-width of graphs defined by one-vertex extensions". In: *Discrete Mathematics* 308.24 (2008), pp. 6157–6165.

[28] M. Rao. "Last cases of Dejean's conjecture". In: *Theoretical Computer Science* 412.27 (2011), pp. 3010–3018.

[29] M. Rao. "MSOL partitioning problems on graphs of bounded treewidth and clique-width". In: *Theoretical Computer Science* 377.1 (2007), pp. 260–267.

[30] M. Rao. "Solving some NP-complete problems using split decomposition". In: *Discrete Applied Mathematics* 156.14 (2008), pp. 2768–2780.

[31] M. Rao, M. Rigo and P. Salimov. "Avoiding 2-binomial squares and cubes". In: *Theoretical Computer Science* 572 (2015), pp. 83–91.

[32] M. Rao and E. Vaslet. "Dejean words with frequency constraint". In: *Manuscript* (2013).

Références

[33] S. O. Aanderaa and H. R. Lewis. "Linear Sampling and the ∀∃∀ Case of the Decision Problem". In: *Journal of Symbolic Logic* 39.3 (Sept. 1974). See also [110], pp. 519–548.

[34] A. Aberkane, J. D. Currie and N. Rampersad. "The number of ternary words avoiding abelian cubes grows exponentially". In: *J. Integer Sequences* 7.2 (2004), Art. 04.2.7.

[35] C. Allauzen and B. Durand. "Tiling Problems". In: *The Classical Decision Problem*. Perspectives in Mathematical Logic. Springer, 2001. Chap. A, pp. 407–420.

[36] J.-P. Allouche, A. Arnold, J. Berstel, S. Brlek, W. Jockusch, S. Plouffe and B. E. Sagan. "A relative of the Thue-Morse sequence". In: *Discrete Mathematics* 139.1–3 (1995), pp. 455–461.

[37] G. Badkobeh. "Fewest repetitions vs maximal-exponent powers in infinite binary words". In: *Theoret. Comput. Sci.* 412.48 (2011), pp. 6625–6633.

[38] G. Badkobeh and M. Crochemore. "Fewest repetitions in infinite binary words". In: *RAIRO - Theor. Inf. and Applic.* 46.1 (2012), pp. 17–31.

[39] G. Badkobeh and M. Crochemore. "Finite-Repetition threshold for infinite ternary words". In: *WORDS*. 2011, pp. 37–43.

[40] R. Berger. "The Undecidability of the Domino Problem". PhD thesis. Harvard University, 1964.

[41] R. Berger. *The Undecidability of the Domino Problem*. Memoirs of the American Mathematical Society 66. The American Mathematical Society, 1966.

[42] J. Berstel. "Axel Thue's papers on repetitions in words: a translation". In: *Monographies du LaCIM*. Vol. 11. incollection. LaCIM, 1992, pp. 65–80.

[43] J. Berstel and P. Séébold. "Sturmian Words". In: *Algebraic Combinatorics on Words*. Cambridge University Press, 2002. Chap. 2, pp. 40–97.

[44] O. Borodin, A. Glebov, A. Raspaud and M. Salavatipour. "Planar graphs without cycles of length from 4 to 7 are 3-colorable". In: *Journal of Combinatorial Theory, Series B* 93.2 (2005), pp. 303–311.

[45] F.-J. Brandenburg. "Uniformly growing k-th power-free homomorphisms". In: *Theoretical Computer Science* 23.1 (1988), pp. 69–82.

[46] A. Carpi. "On Abelian Power-Free Morphisms". In: *International Journal of Algebra and Computation* 03.02 (1993), pp. 151–167.

[47] A. Carpi. "On Abelian squares and substitutions". In: *Theoretical Computer Science* 218.1 (1999), pp. 61–81.

[48] A. Carpi. "On the number of Abelian square-free words on four letters". In: *Disc. Appl. Math* 81.1-3 (1998), pp. 155–167.

[49] A. Carpi. "On Dejean's conjecture over large alphabets". In: *Theoret. Comput. Sci.* 385.1-3 (2007), pp. 137–151.

[50] J. Cassaigne, J. D. Currie, L. Schaeffer and J. Shallit. "Avoiding Three Consecutive Blocks of the Same Size and Same Sum". In: *Journal of the ACM* 61.2 (2014), 10:1–10:17.

[51] J. Chalopin and P. Ochem. "Dejean's conjecture and letter frequency". In: *RAIRO-Theor. Inf. Appl.* 42.3 (2008), pp. 477–480.

[52] T. Y. Chang and W. E. Clark. "The domination numbers of the 5 × n and 6 × n grid graphs". In: *Journal of Graph Theory* 17.1 (1993), pp. 81–107.

[53] T. Chang. "Domination Numbers of Grid Graphs". PhD thesis. University of South Florida, 1992.

[54] T. Chang, W. Clark and E. Hare. "Dominations of complete grid graphs I". In: *Ars Combin.* 38 (1994), pp. 97–112.

[55] H.-H. Chen, W.-G. Hu, D.-J. Lai and S.-S. Lin. "Nonemptiness problems of Wang tiles with three colors". In: *Theoretical Computer Science* 547 (Aug. 2014), pp. 34–45.

[56] E. Cockayne, E. Hare, S. Hedetniemi and T. Wimer. "Bounds for the Domination Number of Grid Graphs". In: *Congressus Numeratium* 47 (1985), pp. 217–228.

[57] V. Cohen-Addad, M. Hebdige, D. Král', Z. Li and E. Salgado. "Steinberg's Conjecture is false". In: *Journal of Combinatorial Theory, Series B* 122 (2017), pp. 452–456.

[58] K. Culik II. "An aperiodic set of 13 Wang tiles". In: *Discrete Mathematics* 160 (1996), pp. 245–251.

[59] J. D. Currie and N. Rampersad. "Fixed Points Avoiding Abelian K-powers". In: *Journal of Combinatorial Theory, Series A* 119.5 (July 2012), pp. 942–948.

[60] J. D. Currie. "The number of binary words avoiding abelian fourth powers grows exponentially". In: *Theoretical Computer Science* 319.1 (2004), pp. 441–446.

[61] J. D. Currie and N. Rampersad. "A proof of Dejean's conjecture". In: *Math. Comput.* 80.274 (2011), pp. 1063–1070.

[62] J. D. Currie and N. Rampersad. "Dejean's conjecture holds for $n \geq 30$". In: *Theoret. Comput. Sci.* 410.30-32 (2009), pp. 2885–2888.

[63] J. Currie and N. Rampersad. "Dejean's conjecture holds for N ≥ 27". In: *RAIRO-Theor. Inf. Appl.* 43.4 (2009), pp. 775–778.

[64] F. Dejean. "Sur un théorème de Thue". In: *Journal of Combinatorial Theory, Series A* 13.1 (1972), pp. 90–99.

[65] F. M. Dekking. "Strongly non-repetitive sequences and progression-free sets". In: *Journal of Combinatorial Theory, Series A* 27.2 (1979), pp. 181–185.

[66] B. Durand, L. A. Levin and A. Shen. "Local rules and global order, or aperiodic tilings". In: *Mathematical Intelligencer* 27.1 (2004), pp. 64–68.

[67] B. Durand, A. Romashchenko and A. Shen. "Fixed-point tile sets and their applications". In: *Journal of Computer and System Sciences* 78.3 (May 2012), pp. 731–764.

[68] B. Durand, A. Shen and A. Romashchenko. *Fixed Point and Aperiodic Tilings*. Tech. rep. TR08-030. ECCC, 2008.

[69] A. E. Edlin. "The number of binary cube-free words of length up to 47 and their numerical analysis". In: *Journal of Difference Equations and Applications* 5.4-5 (1999), pp. 353–354.

[70] R. C. Entringer, D. E. Jackson and J. A. Schatz. "On nonrepetitive sequences". In: *Journal of Combinatorial Theory, Series A* 16.2 (1974), pp. 159–164.

[71] P. Erdős. "Some unsolved problems." In: *The Michigan Mathematical Journal* 4.3 (1957), pp. 291–300.

[72] P. Erdős. "Some unsolved problems". In: *Magyar Tud. Akad. Mat. Kutató Int. Közl.* 6 (1961), pp. 221–254.

[73] A. A. Evdokimov. "Strongly asymmetric sequences generated by a finite number of symbols." In: *Dokl. Akad. Nauk SSSR* 179 (1968), pp. 1268–1271.

[74] T. Fernique and N. Ollinger. "Combinatorial Substitutions and Sofic Tilings ". In: *Journées Automates Cellulaires (JAC)*. TUCS. 2010, pp. 100–110.

[75] F. Fiorenzi, P. Ochem and E. Vaslet. "Bounds for the generalized repetition threshold". In: *Theoretical Computer Science* 412.27 (2011), pp. 2955–2963.

[76] D. C. Fisher. "The Domination number of complete grid graphs". In: *Manuscript* (1993).

[77] A. S. Fraenkel and R. J. Simpson. "How Many Squares Must a Binary Sequence Contain?" In: *The Electronic Journal of Combinatorics* 2 (1995).

[78] A. J. Franz Gähler and J. Savinien. "Combinatorics and Topology of the Robinson tiling". In: *Comptes Rendus de l'Académie des Sciences de Paris, Série I - Mathématiques* (2012), pp. 627–631.

[79] B. le Gloannec and N. Ollinger. "Substitutions and Strongly Deterministic Tilesets". In: *Computability in Europe (CiE)*. Lecture Notes in Computer Science 7318. Springer, 2012, pp. 462–471.

[80] C. Goodman-Strauss. "Matching Rules and Substitution Tilings". In: *Annals of Mathematics* 147.1 (1998), pp. 181–223.

[81] B. Grunbaum and G. Shephard. *Tilings and Patterns*. W.H. Freeman, 1987.

[82] D. R. Guichard. “A lower bound for the domination number of complete grid graphs”. In: *J. Combin. Math. Combin. Comput.* 49 (2004), pp. 215–220.

[83] E. Hare, S. Hedetniemi and W. Hare. “Algorithms for Computing the Domination Number of $K \times N$ Complete Grid Graphs”. In: *Congressus Numeratium* 55 (1986), pp. 81–92.

[84] H. Hermes. “A simplified proof for the unsolvability of the decision problem in the case $\wedge \vee \wedge$”. In: *Logic Colloquium 1969*. Vol. 61. Studies in Logic and the Foundations of Mathematics. 1971, pp. 307–310.

[85] H. Hermes. “Entscheidungsproblem und Dominospiele”. In: *Selecta Mathematica II*. Springer, 1970.

[86] J. E. Hopcroft. *An $n \log n$ algorithm for minimizing states in a finite automaton*. Tech. rep. STAN-CS-71-190. Stanford university, computer science department, Jan. 1971.

[87] M. Huova. “Existence of an infinite ternary 64-abelian square-free word”. In: *RAIRO-Theor. Inf. Appl.* 48.3 (2014), pp. 307–314.

[88] M. Huova, J. Karhumäki and A. Saarela. “Problems in between words and abelian words: k-abelian avoidability”. In: *Theoretical Computer Science* 454 (2012), pp. 172–177.

[89] M. Huova, J. Karhumäki, A. Saarela and K. Saari. “Local Squares, Periodicity and Finite Automata”. In: *Rainbow of Computer Science: Dedicated to Hermann Maurer on the Occasion of His 70th Birthday*. Ed. by C. S. Calude, G. Rozenberg and A. Salomaa. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 90–101.

[90] L. Ilie, P. Ochem and J. Shallit. “A generalization of repetition threshold”. In: *Theoretical Computer Science* 345.2 (2005), pp. 359–369.

[91] M. S. Jacobson and L. F. Kinch. “On the domination number of products of a graph; I”. In: *Ars Comb.* 10 (1983), pp. 33–44.

[92] E. Jeandel and N. Rolin. “Fixed Parameter Undecidability for Wang Tilesets”. In: *Symposium on Cellular Automata (JAC)*. 2012, pp. 69–85.

[93] A. Johnson and K. Madden. “Putting the Pieces Together: Understanding Robinson’s Nonperiodic Tilings”. In: *The College Mathematics Journal* 28.3 (May 1997), pp. 172–181.

[94] J. Justin. “Généralisation du théorème de van der waerden sur les semigroupes répétitifs”. In: *Journal of Combinatorial Theory, Series A* 12.3 (1972), pp. 357–367.

[95] A. Kahr, E. F. Moore and H. Wang. “Entscheidungsproblem reduced to the $\forall\exists\forall$ case”. In: *Proceedings of the National Academy of Sciences of the United States of America* 48.3 (Mar. 1962), pp. 365–377.

[96] P. C. Kanellakis and S. A. Smolka. “CCS Expressions, Finite State Processes and Three Problems of Equivalence”. In: *Information and Computation* 86 (1990), pp. 43–68.

[97] J. Karhumaki, A. Saarela and L. Q. Zamboni. “On a generalization of Abelian equivalence and complexity of infinite words”. In: *Journal of Combinatorial Theory, Series A* 120.8 (2013), pp. 2189–2206.

[98] J. Karhumäki. “Generalized Parikh mappings and homomorphisms”. In: *Information and Control* 47.3 (1980), pp. 155–165.

[99] J. Karhumäki and J. Shallit. “Polynomial versus exponential growth in repetition-free binary words”. In: *J. Comb. Theory, Ser. A* 105.2 (2004), pp. 335–347.

[100] J. Kari. “A small aperiodic set of Wang tiles”. In: *Discrete Mathematics* 160 (1996), pp. 259–264.

[101] J. Kari. “The Nilpotency Problem of One-Dimensional Cellular Automata”. In: *SIAM Journal on Computing* 21.3 (1992), pp. 571–586.

[102] J. Kari and P. Papasoglu. “Deterministic Aperiodic Tile Sets”. In: *Geometric And Functional Analysis* 9 (1999), pp. 353–369.

[103] V. Keränen. “Abelian Squares are Avoidable on 4 Letters”. In: *ICALP*. 1992, pp. 41–52.

[104] V. Keränen. “New Abelian Square-Free DT0L-Languages over 4 Letters.” In: *Manuscript* (2003).

[105] V. Keränen. “A powerful abelian square-free substitution over 4 letters”. In: *Theoretical Computer Science* 410.38 (2009), pp. 3893–3900.

[106] A. Khalyavin. “The Minimal Density of a Letter in an Infinite Ternary Square-Free Word is 883/3215”. In: *J. Integer Sequences* 10 (2007), Art. 07.6.5.

[107] D. E. Knuth. “Fundamental Algorithms”. In: First. Vol. 1. The Art of Computer Programming. Addison-Wesley, 1968. Chap. 2.3.4.3 The infinity lemma, pp. 382–385.

[108] R. Kolpakov. “Efficient Lower Bounds on the Number of Repetition-free Words”. In: *J. Integer Sequences* 10 (2007), Art. 07.3.2.

[109] L. A. Levin. “Aperiodic Tilings: Breaking Translational Symmetry”. In: *The Computer Journal* 48.6 (2005), pp. 642–645.

[110] H. R. Lewis. *Unsolvable Classes of Quantificational Formulas.* Addison-Wesley, 1979.

[111] M. Lothaire, ed. *Combinatorics on Words.* Second. Cambridge University Press, 1997.

[112] Lucian Ilie and Sheng Yu. “Reducing NFAs by invariant equivalences”. In: *Theoretical Computer Science* 306 (2003), pp. 373–390.

[113] V. Lukkarila. “The 4-way deterministic tiling problem is undecidable”. In: *Theoretical Computer Science* 410 (2009), pp. 1516–1533.

[114] J. K. M. Huova. “On the Unavoidability of k-Abelian Squares in Pure Morphic Words”. In: *J. Integer Sequences* 16 (2013), Art. 13.2.9.

[115] R. Mercaş and A. Saarela. “3-Abelian Cubes Are Avoidable on Binary Alphabets”. In: *Developments in Language Theory: 17th International Conference, DLT 2013, Marne-la-Vallée, France, June 18-21, 2013. Proceedings.* Ed. by M.-P. Béal and O. Carton. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 374–383.

[116] R. Mercaş and A. Saarela. “5-Abelian cubes are avoidable on binary alphabets”. In: *RAIRO-Theor. Inf. Appl.* 48.4 (2014), pp. 467–478.

[117] M. Mohammad-Noori and J. D. Currie. “Dejean’s conjecture and Sturmian words”. In: *European J. Combin.* 28.3 (2007), pp. 876–890.

[118] J. Moulin Ollagnier. “Proof of Dejean’s conjecture for alphabets with $5, 6, 7, 8, 9, 10$ and 11 letters”. In: *Theoret. Comput. Sci.* 95.2 (1992), pp. 187–205.

[119] S. Mozes. “Tilings, substitutions systems and dynamical systems generated by them”. In: *J. d’Analyse Math.* 53 (1989), pp. 139–186.

[120] P. Ochem. “A generator of morphisms for infinite words”. In: *RAIRO - Theoret. Informatics Appl.* 40 (2006), pp. 427–441.

[121] P. Ochem and T. Reix. “Upper bound on the number of ternary square-free words”. In: *Workshop on Words and Automata*. 2006.

[122] P. Ochem. “Letter frequency in infinite repetition-free words”. In: *Theoretical Computer Science* 380.3 (2007), pp. 388–392.

[123] N. Ollinger. “Two-by-Two Substitution Systems and the Undecidability of the Domino Problem”. In: *CiE 2008*. Lecture Notes in Computer Science 5028. 2008, pp. 476–485.

[124] R. Paige and R. E. Tarjan. “Three Partition Refinement Algorithms”. In: *SIAM Journal on Computing* 16.6 (Dec. 1987), pp. 973–989.

[125] J.-J. Pansiot. “À propos d’une conjecture de F. Dejean sur les répétitions dans les mots”. In: *Discrete Appl. Math.* 7.3 (1984), pp. 297–311.

[126] G. Pirillo and S. Varricchio. “On uniformly repetitive semigroups”. English. In: *Semigroup Forum* 49.1 (1994), pp. 125–129.

[127] P. A. B. Pleasants. “Non-repetitive sequences”. In: *Mathematical Proceedings of the Cambridge Philosophical Society* 68 (02 Sept. 1970), pp. 267–274.

[128] B. Poizat. “Une théorie finiement axiomatisable et superstable”. In: *Groupe d’études de théories stables* 3 (1980), pp. 1–9.

[129] V. Poupet. “Yet Another Aperiodic Tile Set ”. In: *Journées Automates Cellulaires (JAC)*. TUCS. 2010, pp. 191–202.

[130] R. Kolpakov, G. Kucherov and Y. Tarannikov. “On repetition-free binary words of minimal density”. In: *Theoret. Comput. Sci.* 218 (1999), pp. 161–175.

[131] N. Rampersad and J. Shallit. “Repetitions in words”. In: *Combinatorics, Words and Symbolic Dynamics*. Ed. by V. Berthé and M. Rigo. Cambridge University Press, 2016, pp. 101–150.

[132] N. Rampersad, J. Shallit and M.-W. Wang. “Avoiding large squares in infinite binary words”. In: *Theor. Comput. Sci.* 339.1 (2005), pp. 19–34.

[133] M. Rigo and P. Salimov. “Another Generalization of Abelian Equivalence: Binomial Complexity of Infinite Words”. In: *Combinatorics on Words: 9th International Conference, WORDS 2013, Turku, Finland, September 16-20. Proceedings*. Ed. by J. Karhumäki, A. Lepistö and L. Zamboni. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 217–228.

[134] R. M. Robinson. “Undecidability and Nonperiodicity for Tilings of the Plane”. In: *Inventiones Mathematicae* 12.3 (1971), pp. 177–209.

[135] M. Rosenfeld. "Avoidability of abelian repetitions in words". PhD thesis. École Normale Supérieure de Lyon, 2017.

[136] O. Salon. "Quelles tuiles! (Pavages apériodiques du plan et automates bidimensionnels)". In: *Journal de Théorie des Nombres de Bordeaux* 1.1 (1989), pp. 1–26.

[137] M. Senechal. *Quasicrystals and geometry*. Cambridge University Press, 1995.

[138] J. Shallit. "Simultaneous avoidance of large squares and fractional powers in infinite binary words". In: *Intl. J. Found. Comput. Sci* 15.2 (2004), pp. 317–327.

[139] A. M. Shur. "Growth rates of complexity of power-free languages". In: *Theoretical Computer Science* 411.34 (2010), pp. 3209–3223.

[140] A. M. Shur and I. A. Gorbunova. "On the growth rates of complexity of threshold languages". In: *RAIRO-Theor. Inf. Appl.* 44.1 (2010), pp. 175–192.

[141] H. Singh and R. Pargas. "A parallel implementation for the domination number of grid graphs". In: *Cong. Numer.* 59 (1987).

[142] A. Thue. "Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen." In: *Norske Vid. Selsk. Skr. I. Mat. Nat. Kl. Christiania,* 10 (1912), pp. 1–67.

[143] A. Thue. "Über unendliche Zeichenreihen". In: *Norske Vid. Selsk. Skr. I. Mat. Nat. Kl. Christiania,* 7 (1906), pp. 1–22.

[144] I. N. Tunev and A. M. Shur. "On Two Stronger Versions of Dejean's Conjecture". In: *Mathematical Foundations of Computer Science 2012: 37th International Symposium, MFCS 2012, Bratislava, Slovakia, August 27-31, 2012. Proceedings.* Ed. by B. Rovan, V. Sassone and P. Widmayer. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 800–812.

[145] A. Valmari. "Simple Bisimilarity Minimization in $\mathrm{O}(m \log n)$ Time". In: *Fundamenta Informaticae* 105 (2010), pp. 319–339.

[146] H. Wang. "Notes on a class of tiling problems". In: *Fundamenta Mathematicae* 82 (1975), pp. 295–305.

[147] H. Wang. "Proving theorems by Pattern Recognition II". In: *Bell Systems technical journal* 40 (1961), pp. 1–41.

[148] Z.-X. Wen and Z.-Y. Wen. "Some Properties of the Singular Words of the Fibonacci Word". In: *European Journal of Combinatorics* 15.6 (Nov. 1994), pp. 587–598.