

# Robustesse des réseaux d'automates booléens à seuil aux modes d'itération. Application à la modélisation des réseaux de régulation génétique.

Adrien Elena

▶ **To cite this version:**

Adrien Elena. Robustesse des réseaux d'automates booléens à seuil aux modes d'itération. Application à la modélisation des réseaux de régulation génétique.. Modélisation et simulation. Université Joseph-Fourier - Grenoble I, 2009. Français. NNT : . tel-00447564

**HAL Id: tel-00447564**
**https://theses.hal.science/tel-00447564**

Submitted on 15 Jan 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Joseph Fourier (Grenoble I)

Laboratoire des Techniques de l'Ingénierie Médicale et de la Complexité - Informatique, Mathématiques et Applications de Grenoble

THÈSE

*présentée et soutenue publiquement le 12 mai 2009 par*

**Adrien Elena**

*pour l'obtention du grade de*

**Docteur de l'Université Joseph Fourier (Grenoble I)**

**spécialité : Modèles, méthodes et algorithmes en biologie, santé et environnement**

*au titre de l'École Doctorale Ingénierie pour la Santé, la Cognition et l'Environnement de Grenoble*

# Robustesse des réseaux d'automates booléens à seuil aux modes d'itération. Application à la modélisation des réseaux de régulation génétique.

Directeur de thèse

Jacques Demongeot

Jury

| | | |
|---|---|---|
| M. | Julio Aracena | Examinateur |
| M. | Jean-Paul Comet | Examinateur |
| M. | Jacques Demongeot | Directeur |
| M. | Olivier François | Examinateur |
| M. | Eric Goles | Rapporteur |
| M. | Michel Morvan | Rapporteur |

*À ma famille,*
*À Émilie.*

*À la mémoire de mon grand-père, Jacques, et de mon ami, Loïc.*

# Remerciements

*Je tiens en tout premier lieu à remercier Jacques Demongeot pour m'avoir proposé un thème de recherche aussi intéressant, et pour avoir toujours été d'une écoute et d'un soutien inestimables depuis notre première rencontre. Malgré un "emploi du temps de ministre", il a su être présent, et sa passion, son immense gentillesse m'ont grandement aidé à mener à bien cette thèse.*

*Un grand merci à mes deux rapporteurs, Eric Goles et Michel Morvan, qui ont accepté de lire cette thèse en détail afin d'en juger la qualité. Je voudrais également remercier Julio Aracena et Jean-Paul Comet pour leur participation, en tant qu'examinateurs, à ma soutenance. Je souhaite adresser un remerciement spécial à Olivier François qui m'a fait l'honneur d'être président du jury. Ce fut un président particulier pour moi, puisqu'en tant qu'enseignant, c'est lui qui m'a introduit dans le monde de la recherche. Il est donc, d'une certaine manière, à l'origine de cette thèse. Je remercie tous les membres du jury de ma thèse pour leurs remarques et leurs questions pertinentes sur mes travaux.*

*Je tiens également à remercier tous les membres de l'équipe TIMB pour leur accueil, leurs encouragements et leur soutien. Merci donc à Jean-Louis, Olivier, Hervé, Cécile, Eric, Nicolas et Laurent. Bien entendu, je n'oublie pas les "jeunes", sans qui ces années de thésard n'auraient pas la même saveur. Un grand merci à Hedi, Loïc, Mathieu, Michael, Nicolas et Sylvain pour toutes les discussions plus ou moins sérieuses et tous les moments de détente.*

*Je souhaite témoigner toute ma gratitude à Céline Fontant, Caroline Messina-Dos Santos, Nadine Perraud et Kim Pla, assistantes de Jacques Demongeot. Elles ont participé au succès de cette thèse.*

*Je ne remercierai jamais assez mes parents et mon frère, pour tout ce qu'ils ont fait pendant cette thèse, mais également avant. D'un soutien sans faille, ils m'ont accompagné, répondant toujours présents lorsque j'avais besoin d'eux. Je tiens à remercier également ma famille, mes amis qui m'ont suivi et ont toujours montré de l'intérêt à ce que je faisais pendant ces années. Enfin, et avant tout, mes pensées vont à Emilie, que je remercie pour tout le bonheur que représentent ces années à mes côtés.*

# Table des matières

**Table des matières** **vii**

**Notations** **ix**

**Introduction** **1**

**1 Préliminaires** **9**
1.1 Systèmes dynamiques discrets . . . . . 10
1.2 Réseaux d'automates et réseaux de régulation biologique . . . . . 13
1.2.1 Réseaux d'automates . . . . . 14
1.2.2 Modèles de réseaux de régulation biologique . . . . . 17
1.3 Modes d'itération et robustesse . . . . . 27
1.3.1 Modes d'itération . . . . . 28
1.3.2 Robustesse . . . . . 31

**2 Enumérations et algorithme pour la simulation** **35**
2.1 A propos des permutations . . . . . 37
2.2 Enumération des fonctions booléennes à seuil . . . . . 39
2.3 Enumération des réseaux d'automates booléens à seuil . . . . . 51
2.4 Algorithme pour la simulation des réseaux . . . . . 62

**3 Résultats expérimentaux** **69**
3.1 Statistiques générales . . . . . 70
3.1.1 Classes de comportement dynamique . . . . . 70
3.1.2 Attracteurs . . . . . 81
3.2 Etude de la classe de comportement *Ev* . . . . . 84
3.3 L'exemple de la morphogénèse florale chez *Arabidopsis thaliana* . 92
3.4 *miARN* et robustesse . . . . . 102

**4 Itération bio-inspirée** **107**
4.1 Motivation et premières définitions . . . . . 107
4.2 Résultats . . . . . 114

**Conclusion et perspectives** **125**

**Bibliographie** **137**

# Notations

Vous trouverez ci-desous une liste des notations les plus fréquemment utilisées tout au long de ce document

- $\mathbb{R}$ représente l'ensemble des nombre réels ($\mathbb{R}^*$ est égal à $\mathbb{R}$ privé de 0) ;
- $\mathbb{Q}$ représente l'ensemble des nombres rationnels ($\mathbb{Q}^*$ est égal à $\mathbb{Q}$ privé de 0) ;
- $\mathbb{Z}$ représente l'ensemble des nombres entiers relatifs ($\mathbb{Z}^*$ est égal à $\mathbb{Z}$ privé de 0) ;
- $\mathbb{N}$ représente l'ensemble des nombres entier naturels ($\mathbb{N}^*$ est égal à $\mathbb{N}$ privé de 0) ;
- sauf mention particulière, $n \in \mathbb{N}^*$, c'est-à-dire, $n$ est un entier naturel non nul ;
- si $E$ est un ensemble, alors $\#E$ représente le cardinal de $E$, c'est-à-dire le nombre d'éléments de $E$ ;
- $\{a, \cdots, b\}$ avec $(a, b) \in \mathbb{Z}^2$ représente l'ensemble des entiers relatifs compris entre $a$ et $b$ ;
- $X$ et $\mathbf{X}$ sont à différencier : $X$ peut représenter un scalaire (c'est le cas la plupart du temps), alors que $\mathbf{X}$ (écrit en gras) représente un objet à plusieurs dimensions, généralement un vecteur ou une matrice ;
- si $\mathbf{X}$ est un vecteur de $\mathbb{R}^n$, alors, sauf mention particulière, $\mathbf{X} = (x_1, \cdots, x_n)$ ; les coordonnées d'un vecteur sont désignées par la même lettre (en minuscule) que le vecteur lui-même ;

# Introduction

## Contexte

En 1865, Mendel pose les bases théoriques de la génétique moderne. A partir de ses études expérimentales menées sur les pois, il énonce les lois (théoriques) de l'hérédité [1] : certains caractères morphologiques sont transmis à travers plusieurs générations. Le terme de "gène" n'est pas encore employé, et le support moléculaire de l'information transmise n'est pas connu, mais l'existence de facteurs biologiques de l'hérédité est désormais mise en évidence. Les travaux de Mendel ne rencontrent pas un accueil favorable auprès de ses contemporains. Il faut attendre 1889 pour que De Vries [2] redécouvre expérimentalement les lois de Mendel. De Vries reprend alors le terme de "pangène", introduit par Darwin. Les "pangènes" seraient des organites [1] intracellulaires, présents dans toutes les cellules. En 1909, Johannsen est le premier à utiliser le terme de "gène" (contraction de "pangène"). Dès lors, les facteurs d'hérédité introduits par Mendel sont décrits par les gènes. Johannsen se démarque cependant de De Vries. Il considère que le gène doit être dégagé de toute interprétation morphologique particulière, et propose de le définir de manière purement opérationnelle : "Il faut traiter le gène comme une unité de comptage ou de calcul. Nous n'avons aucunement le droit de définir le gène comme une structure morphologique" [3]. Johannsen introduit également les notions de génotype et de phénotype : le génotype est l'ensemble (ou une partie donnée) de la combinaison génétique (information génétique) d'un individu ; le phénotype est l'ensemble (ou une partie donnée) des caractères observables caractérisant un individu. Le gène est une unité de fonction, qui n'est révélée comme telle que par le phénotype correspondant à une combinaison génotypique donnée. Cependant, dès 1910, la génétique devient chromosomique : les études expérimentales de Morgan sur la mouche drosophile montrent que certains caractères héréditaires sont liés au sexe, et que ces caractères sont probablement portés par l'un des chromosomes sexuels. Morgan en déduit que les autres gènes sont également portés par des chromosomes spécifiques [4]. Cette découverte vaudra le prix Nobel de médecine à Morgan, en 1933. En 1941, Beadle et Tatum découvrent le lien existant entre un gène et une enzyme [5]. Ils exposent de la moisissure de pain à des rayons X, et provoquent ainsi des mutations génétiques. Leurs expériences montrent alors que ces mutations ont pour effet de

[1] Les organites sont des structures cellulaires spécialisées, caractéristiques des cellules eucaryotes, situées dans le cytoplasme et délimitées par une membrane.

modifier des enzymes spécifiques : il existerait donc un lien direct entre les gènes et les enzymes. On parle de l'hypothèse "un gène, une enzyme" : chaque gène a pour fonction de diriger la production d'une enzyme particulière. Les gènes déterminent les phénotypes par l'intermédiaire d'enzymes catalysant certaines réactions chimiques. Beadle et Tatum reçoivent le prix Nobel de médecine en 1958, pour leur découverte de la relation "un gène, une enzyme", qui se généralise à "un gène, une protéine" (toutes les protéines ne sont pas des enzymes). A la fin des années 1940, il est communément admis que l'ADN (acide désoxyribonucléique) est le vecteur de l'hérédité, et qu'il porte donc l'information génétique. En revanche, la structure de cette molécule est encore inconnue. C'est en 1953 que Crick et Watson [6] déterminent la structure fine de l'ADN, sous forme de double-hélice. Cette découverte est primordiale car la connaissance de la molécule constituant les gènes va aider à comprendre les mécanismes moléculaires de l'hérédité. En 1962, le prix Nobel de médecine est décerné à Crick, Watson et Wilkins (qui a collaboré à la découverte de la structure de l'ADN). Dans les années 1960, Jacob et Monod [7] mettent en évidence la molécule qui fait le lien entre le génome et les protéines, l'ARN messager, et découvrent les mécanismes de la régulation génétique. Leurs travaux sont récompensés, en 1965, par le prix Nobel de médecine.

Depuis les travaux de Jacob et Monod, on connaît la façon dont les gènes permettent de synthétiser les protéines. Ce processus, appelé processus d'expression génétique, se déroule en plusieurs étapes :

1. la *transcription*, qui est la synthèse de l'ARN messager à partir de l'ADN ; la molécule d'ARN messager obtenue à partir de l'ADN est une transcription fidèle des instructions fournies par un gène, en vue de la synthèse d'une protéine ;
2. la phase de *maturation* qui permet de modifier l'ARN messager transcrit, afin de ne conserver que les parties de la molécule utiles à la fabrication d'une protéine ;
3. enfin, la *traduction* représente la synthèse de la protéine à partir des instructions contenues dans l'ARN messager.

L'expression génétique est un processus qui est régulé lors des différentes étapes de la synthèse des protéines. La régulation la mieux connue concerne la transcription (la régulation concerne également les phases de maturation et de transport de l'ARN messager). Les travaux récents montrent cependant que la régulation au niveau de la traduction joue également un rôle très important [8]. La transcription d'un gène peut être activée ou, au contraire, inhibée, par la présence d'une molécule (ARN ou protéine) sur le site promoteur du gène (le site promoteur d'un gène représente la région particulière de l'ADN sur laquelle vient se fixer l'enzyme qui va catalyser la réaction de transcription du dit gène). Ces molécules, activatrices, ou inhibitrices, sont les produits d'autres gènes. Ainsi, sont mis en place des systèmes de régulation génétique sous forme de réseaux d'interaction entre ADN, ARN, protéines et autres petites molécules.

Les protéines sont impliquées dans de très nombreux processus cellulaires,

comme la réponse d'une cellule à une modification de son environnement, la différenciation cellulaire ou encore la réplication de l'ADN. Si l'on veut, un jour, comprendre le vivant, au niveau de la cellule, il nous faut comprendre la façon dont les protéines sont synthétisées, c'est-à-dire la façon dont les gènes sont exprimés. En d'autres termes, pour comprendre le fonctionnement des être vivants, au niveau moléculaire, nous devons comprendre quels sont les gènes qui sont exprimés, à quel moment et où, dans l'organisme. Les techniques de biologie moléculaire modernes permettent d'acquérir une masse énorme de données expérimentales concernant l'expression génétique. On peut, par exemple, mesurer en parallèle l'expression de milliers de gènes grâce aux bio-puces [9, 10, 11, 12]. Ces données sont censées nous permettre de mieux comprendre le vivant. Mais il nous faut tout d'abord franchir l'obstacle constitué par les réseaux de régulation génétique. Si la description du processus d'expression d'un gène peut se faire relativement simplement, en revanche, décrire l'évolution simultanée de l'expression de l'ensemble des gènes d'un réseau de régulation reste une tâche ardue. En ce sens, les réseaux de régulation génétique font partie d'une classe de systèmes dynamiques, que l'on appelle "systèmes complexes".

Il n'existe pas de définition formelle de ce qu'est un système complexe, mais nous pouvons cependant le caractériser de la façon suivante : un système complexe est composé d'un ensemble d'éléments en interaction, et est tel que le comportement dynamique global du système ne peut être expliqué de façon directe, simplement à partir de la connaissance du comportement de chacun des éléments, pris séparément. On trouve des exemple de systèmes complexes dans de nombreuses disciplines, ce qui explique l'intérêt grandissant porté à cette classe de systèmes dynamiques. Nous pouvons citer quelques exemple issus de différentes disciplines :

- en informatique, les réseaux pair-à-pair qui permettent le partage de ressources ;
- en économie, l'ensemble des agents qui effectuent des transactions sur le marché boursier, et dont les décisions peuvent provoquer des phénomènes globaux comme les bulles ou les krachs ;
- en biologie, les colonies de fourmis, qui montrent que des individus au comportement primaire peuvent effectuer collectivement une tâche complexe de façon efficace, comme la collecte de nourriture.

Afin d'étudier le comportement global des systèmes complexes, une phase de modélisation est nécessaire. Elle est source de travaux interdisciplinaires puisqu'elle doit permettre de répondre à des questions propres à la discipline dont le système est issu, au moyen de modèles définis dans un formalisme appartenant à une autre discipline. Dans le cadre de notre travail, afin de comprendre comment les gènes d'un réseau de régulation sont exprimés, nous allons utiliser un modèle mathématique. Ce travail de doctorat s'inscrit dans la volonté d'apporter des réponses à certaines questions autour de systèmes complexes biologiques tels que les réseaux de régulation génétique, à partir d'une étude réalisée dans le cadre formel des mathématiques discrètes.

## Problématique

L'objet de notre travail n'est pas, à proprement parler, la modélisation des systèmes de régulation génétique, bien qu'il s'inscrive clairement dans cette thématique, mais concerne plutôt l'étude d'un modèle particulier de réseaux de régulation génétique. Nous avons vu que les réseaux de régulation génétique, en tant que systèmes complexes, doivent être modélisés afin d'être étudiés. Le but d'un modèle est de décrire la réalité, biologique dans notre cas. Il ne peut pas prendre en compte tous les mécanismes existants et doit justement mettre en exergue les points importants permettant d'expliquer ce que l'on observe. Dans la pratique, on met au point un modèle à partir d'une connaissance existante, dans le but de décrire puis de comprendre les mécanismes sur lesquels porte l'étude. Une fois le modèle mis au point, il faut l'éprouver, en comparant par exemple les prévisions issues de simulations réalisées à partir du modèle et les données expérimentales, afin de vérifier qu'il y a adéquation entre ce qui est prédit par le modèle et ce que l'on observe. Tant qu'il n'y a pas adéquation, il faut reprendre, améliorer le modèle. Le but ultime d'un modèle est de prédire ce qui peut se passer à partir d'une configuration initiale donnée du système étudié. Mais avant d'atteindre cette étape de prévision à partir du modèle, il faut comprendre parfaitement le modèle en tant qu'objet à part entière, afin de s'assurer que le comportement que l'on obtient par simulations à partir du modèle, décrit fidèlement la réalité du système étudié. Si ce n'est pas le cas, il existe un risque de vouloir interpréter un comportement obtenu par simulations, alors que le modèle n'est pas tout à fait adéquat. L'étude du modèle nous permet de comprendre si les caractéristiques obtenues par simulations à partir du modèle sont des propriétés inhérentes au modèle, ou si ces caractéristiques sont aussi valables pour le système modélisé. C'est de cette étude du modèle dont il est question dans ce manuscrit.

La modélisation des réseaux de régulation génétique, et de leur dynamique (la façon dont l'expression des gènes du réseau évolue) a été l'objet de nombreuses études depuis son introduction, à la fin des années 1960 et au début des années 1970, par les travaux de Kauffman et Thomas. Depuis ces premiers modèles, de nombreux formalismes ont été utilisés pour atteindre une meilleure compréhension des phénomènes biologiques liés à la dynamique de tels systèmes. Nous ne souhaitons pas en dresser ici une liste exhaustive et nous nous contentons de citer, parmi les plus classiques, les systèmes d'équations différentielles ordinaires, partielles ou linéaires par morceaux, les réseaux booléens et leurs généralisations discrètes. Le lecteur intéressé pourra se reporter à l'article de De Jong [13], qui propose une synthèse sur les différentes modélisations existantes.

Dans le cadre de notre travail de doctorat, nous avons travaillé sur un modèle booléen, dans lequel les gènes sont représentés par des éléments ne pouvant prendre que 2 valeurs : 0 ou 1. Ceci signifie que, dans ce modèle, on considère, de façon simplifiée, qu'un gène est soit actif, dans le cas où il est exprimé sous la forme d'un ARN messager, puis d'une protéine, soit inactif, dans le cas où il n'est pas exprimé. Cette simplification booléenne peut paraître bien réductrice, puisque l'on imagine que les gènes sont régulés par les concentrations de protéines

qu'ils synthétisent. Cependant, comme nous l'avons déjà évoqué, les méthodes actuelles de la biologie moléculaire permettent de suivre l'évolution de l'expression des gènes, au moyen de bio-puces. Sans entrer dans des détails techniques, le principe des ces bio-puces est que, pour chaque gène dont on veut étudier l'expression, on peut savoir si la concentration en ARN messager correspondant est inférieure ou supérieure à un certain seuil. Si la concentration en ARN messager est inférieure au seuil, alors on peut en déduire que pour la protéine synthétisée à partir de cet ARN messager, la concentration est également faible, trop faible pour avoir un effet en tant que protéine régulatrice. Ainsi, de par l'utilisation des bio-puces pour mesurer l'expression des gènes, les biologistes sont amenés à utiliser la terminologie booléenne, par exemple, avec des phrases du type "si la protéine $pXX$ est présente, le gène $gYY$ s'active (resp. est inhibé)". Pour ces raisons, la modélisation des réseaux de régulation génétique, à l'aide d'un modèle booléen, n'est pas si réductrice qu'il n'y paraît. Elle ne permet certes pas de modéliser le fait qu'une protéine puisse posséder plusieurs niveaux de concentration résultant en différentes régulations d'un même gène. Cependant, une grande partie de la communauté scientifique considère l'approche booléenne comme une approximation raisonnable de la réalité biologique. Les modèles booléens permettent d'améliorer la compréhension de la dynamique des réseaux de régulation génétique, par une description qualitative de cette dynamique.

Dans ce travail, notre choix s'est porté sur une classe particulière de modèles booléens, les réseaux d'automates booléens à seuil. Ces réseaux d'automates ont été introduits, dans les années 1940, pour modéliser les réseaux de neurones, puis ont été réutilisés, dans le cadre de la modélisation des réseaux de régulation génétique, à la fin des années 1990. L'une des principales propriétés des réseaux d'automates est que leur comportement dynamique peut dépendre de l'ordre dans lequel les éléments du réseau voient leur état être mis à jour. Nous rappelons que, dans le cadre de la modélisation des réseaux de régulation génétique, le but du modèle est de décrire l'évolution de l'expression des différents gènes. La mise à jour de l'état d'un élément d'un réseau d'automates modélise donc le processus d'expression du gène représenté par cet élément. Nous étudions l'une des caractéristiques principales des systèmes complexes, et donc des réseaux de régulation génétique, à savoir la robustesse. La robustesse d'un système peut être définie comme sa capacité à résister à des perturbations d'ordre structurel. Ici, la robustesse concerne la capacité du réseau à maintenir son comportement dynamique malgré un changement de l'ordre de mise à jour de ses éléments. Comme pour tous les système complexes, cette notion de robustesse est fondamentale pour les réseaux de régulation génétique. Si les réseaux de régulation génétique ne sont pas robustes, alors l'expression génétique, c'est-à-dire la synthèse des protéines, peut être altérée. Comme nous avons vu que les protéines jouaient un rôle majeur dans de nombreux processus cellulaires, une non robustesse des réseaux de régulation génétique aurait des conséquences importantes sur la vie des organismes considérés.

*Dans cette thèse, nous étudions la robustesse des réseaux d'automates booléens à seuil aux modes d'itération. Nos travaux portent sur l'étude de ces réseaux, en*

*tant qu'objets mathématiques, au moyen de simulations numériques, mais également par des méthodes théoriques. Le but de cette étude est de mieux connaître les réseaux d'automates booléens à seuil, afin de comprendre si leur robustesse (ou leur sensibilité) aux modes d'itération, est une caractéristique qui les rend appropriés à la modélisation des réseaux de régulation génétique. Tout au long du manuscrit, nous considérerons donc ces réseaux d'automates en tant que modèles de réseaux de régulation génétique.*

## Organisation du manuscrit

Le premier chapitre de cette thèse introduit les principales définitions utilisées dans la suite du document. Nous y présentons tout d'abord les systèmes dynamiques discrets et les réseaux d'automates. Nous traitons ensuite, dans la partie 1.2.2, de la modélisation des réseaux de régulation biologique par des réseaux d'automates : nous abordons, dans un premier temps, les modèles proposés par Kauffman et Thomas, puis présentons, de façon détaillée, le modèle que nous étudions, sous forme de réseaux d'automates booléens à seuil, et donnons un état de l'art des résultats concernant les modèles de réseaux de régulation. Enfin, nous terminons ce premier chapitre en précisant, dans la partie 1.3, la notion de robustesse aux modes d'itération, en présentant les principaux résultats théoriques connus concernant la dynamique des réseaux d'automates booléens et la robustesse aux modes d'itération, et en définissant 4 classes de comportement dynamiques pour les réseaux d'automates, qui nous seront très utiles pour l'étude de la robustesse de ces réseaux.

Dans le chapitre 2, nous entrons dans le vif du sujet. Nous y montrons comment on peut énumérer, et donc construire, l'ensemble des réseaux d'automates booléens à seuil, à partir de l'ensemble des fonctions booléennes à seuil. Après avoir rappelé quelques notions sur les permutations dans la partie 2.1, nous introduisons, dans la partie 2.2, les vecteurs d'interaction, qui nous permettent de définir les fonctions booléennes à seuil, en vue de construire les réseaux d'automates booléens à seuil. Nous insistons alors sur l'intérêt d'utiliser des vecteurs d'interaction particuliers, que nous appelons vecteurs d'interaction minimaux, dans le cadre de la modélisation des réseaux de régulation génétique. Dans la partie 2.3, nous donnons l'énumération des réseaux d'automates booléens à seuil, construits à partir de l'énumération des fonctions booléennes à seuil. Cette énumération nous permet de connaître la structure de la population des réseaux, et de donner, dans la partie 2.4, les algorithmes que nous utilisons pour réaliser les simulations dont les résultats sont présentés dans le chapitre 3. Ce chapitre 2 est le chapitre de la transversalité : il montre que, dans le cadre de la modélisation de réseaux biologiques, on peut être amené à utiliser des notions issues de domaines inattendus des mathématiques, telles que la géométrie (pour les fonctions booléennes à seuil) ou la théorie des actions de groupe (pour les réseaux d'automates), en plus des traditionnelles notions de statistique (pour les simulations).

Dans le chapitre 3, nous présentons l'ensemble des résultats des simulations

numériques que nous avons réalisées, concernant la robustesse des réseaux d'automates booléens à seuil. La partie 3.1 donne les statistiques générales en relation avec la robustesse. Nous montrons, par exemple, que, pour les réseaux d'automates à 7 éléments ou moins, l'augmentation de la taille des réseaux se traduit par une baisse de la robustesse de ces réseaux. Nous montrons également que cette baisse de la robustesse s'explique par une connectivité très élevée des réseaux. Dans la partie 3.2, nous étudions, toujours au moyen de simulations, la classe de comportement des réseaux les moins robustes aux perturbations des modes d'itération. Dans la partie 3.3, nous traitons un exemple de modèle de réseau de régulation génétique connu, celui de la floraison d'*Arabidopsis thaliana*, sous forme de réseau d'automates booléens à seuil. Nous montrons, notamment, comment, à partir d'une bonne connaissance du modèle, nous pouvons facilement ramener l'étude fastidieuse de la dynamique d'un réseau à 12 éléments, à l'étude, bien plus simple, de la dynamique de deux réseaux de taille 3. A partir de simulations numériques sur ces deux petits réseaux, nous pouvons donner la dynamique exhaustive du réseau de 12 éléments. Enfin, dans la partie 3.4, nous proposons un protocole de simulations très simple, afin d'étudier l'impact de l'ajout des molécules de *miARN* dans la régulation génétique, en terme de robustesse des réseaux. Notre étude montre que les *miARN* ont tendance à augmenter la robustesse des réseaux.

Dans le dernier chapitre, toujours avec le souci d'étudier les réseaux d'automates booléens à seuil, dans le cadre de la modélisation des réseaux de régulation génétique, nous introduisons une nouvelle façon d'itérer les réseaux d'automates. Cette nouvelle itération a pour but de modéliser au mieux l'évolution de l'expression génétique, telle que nous pouvons l'observer au moyen des bio-puces. A partir de la définition de cette nouvelle itération, nous pouvons démontrer, dans la partie 4.2, un ensemble de résultats théoriques importants concernant la robustesse des réseaux d'automates aux modes d'itération.

# Chapitre 1

# Préliminaires

L'étude des systèmes biologiques, comme celle de tous les systèmes réels, qu'ils soient physiques ou sociaux, passe par une phase de modélisation. Le but de cette modélisation est d'extraire du système réel ses caractéristiques principales, celles qui permettent de le comprendre, d'en expliquer le comportement, ou du moins de mettre en évidence un comportement spécifique. On peut souvent obtenir à partir des systèmes réels un nombre de données aussi important que l'on souhaite et il est nécessaire lors de la modélisation de savoir réduire la taille du problème sous peine de ne rien pouvoir résoudre, et finalement, de ne rien pouvoir expliquer. Pour mettre en place les principes de la gravitation universelle, et expliquer pourquoi la Lune ne tombait pas sur la Terre, Newton n'a pas eu besoin de connaître la nature des sols des différents astres, l'existence ou non d'une atmosphère et le cas échéant, sa composition... A partir de la masse de deux corps ainsi que de la distance qui les sépare, il a pu décrire la force d'attraction mutuelle entre ces deux corps. Il en est de même pour la biologie : il est inutile de vouloir comprendre le vivant en prenant en compte la totalité de l'information disponible. Le modélisateur est amené à faire un choix pour le modèle en fonction de ce qu'il désire mettre en évidence.

L'utilisation de systèmes dynamiques comme modèles de systèmes biologiques n'est pas récente. Dans les années 1920, Volterra [14, 15, 16] a proposé un modèle d'interaction proies-prédateurs sous forme de système dynamique, afin d'expliquer pourquoi la quantité de sardines pêchées en mer Adriatique était plus faible après l'interruption de la pêche due à la guerre, et pourquoi, à la reprise de la pêche, la proportion de requins observés avait augmenté (le système dynamique proposé est donné dans l'exemple 1.1). A la fin des années 1940, Delbrück est l'un des premiers à mettre en évidence la possibilité et l'intérêt de modéliser des systèmes de biologie cellulaire par des systèmes dynamiques. A cette époque, on explique la différenciation cellulaire par l'existence de répliques des gènes, les plasmagènes, dont aurait dépendu le développement des différents tissus. Delbrück propose une autre explication [17] : la différenciation cellulaire pourrait n'être qu'un cas particulier à la biologie d'un phénomène plus général, la multi-stationnarité, propriété des systèmes pouvant atteindre au moins deux états stables, chacun d'eux ayant leur propre bassin d'attraction : "Les passages

des uns aux autres pourraient être, suivant les cas, réversibles ou irréversibles, comme dans les phénomènes de différenciation".

Le présent document traite d'un cas particulier de robustesse des systèmes dynamiques discrets : nous étudions l'influence des modes d'itération sur le comportement des réseaux de régulation biologique. Ce premier chapitre donne des définitions générales concernant les systèmes dynamiques, présente plus en détails les systèmes particuliers étudiés, et enfin décrit la notion de robustesse appliquée aux réseaux de régulation biologique. Nous dressons également un état de l'art des précédents travaux portant sur les systèmes étudiés.

## 1.1 Systèmes dynamiques discrets

Dans le contexte de nos travaux, les systèmes dynamiques sont des objets mathématiques composés d'entités interagissant et évoluant au cours du temps.

**Définition 1.1** (Système dynamique)
*Un* système dynamique *est un triplet* $(S, T, F)$ *où :*

- $S$ *est un sous-ensemble de* $\mathbb{R}^d$*, appelé espace des phases ou espace des états globaux (d représente le nombre d'éléments du système),*
- $T$ *est un groupe (muni de l'opération +), appelé espace des temps,*
- $F$ *est une application,* $F : S \times T \to S$*, appelée fonction de transition globale qui, à chaque état global de* $S$ *et à chaque temps de* $T$*, associe une image dans* $S$*.*

La fonction de transition globale $F$ vérifie :

- $\forall x \in S,\ \ F(x, 0) = x$,
- $\forall x \in S,\ \ \forall (t_1, t_2) \in T^2,\ \ F((F(x, t_1), t_2) = F(x, t_1 + t_2)$.

Elle définit la façon dont l'état global du système évolue au cours du temps, à partir d'un état initial. $F(x, t)$ représente l'état global du système atteint au temps $t$ en partant de l'état global initial (au temps 0) $x$. Notons ici que la définition précédente correspond précisément à un système dynamique réel, c'est-à-dire dont les éléments sont à états dans $\mathbb{R}$. Comme, dans le cadre de ce manuscrit, on ne s'intéresse aux systèmes dynamiques que comme modèle de phénomènes réels, nous nous satisfaisons de cette restriction. Pour les mêmes raisons, les temps sont généralement choisis réels. Dans la suite, on considère donc $T \subseteq \mathbb{R}$.

Dans le cas où la fonction de transition globale est différentiable, on peut représenter le système dynamique par un système d'équations différentielles.

*Exemple* 1.1 *:* Nous avons vu dans l'introduction de ce chapitre que, dans les années 1920, Volterra a proposé un modèle d'interaction proies-prédateurs pour expliquer l'évolution de la taille des populations de proies et de prédateurs. Ce modèle est décrit par un système de 2 équations différentielles, connu sous le nom de système de Lotka-Volterra ($x$ représente la taille de la population des proies, $y$ celle des prédateurs) :

$$\begin{cases} x'(t) = ax(t) - bx(t)y(t) \\ y'(t) = -cy(t) + dx(t)y(t) \end{cases}$$

Les paramètres $a, b, c$ et $d$ sont des réels positifs. En l'absence de prédateur, la population de proies croît de façon exponentielle avec un taux $a$. A l'opposé, en l'absence de proie, la population de prédateurs décroît de façon exponentielle avec un taux $c$. Lorsque proies et prédateurs sont présents, on suppose le taux de prédation proportionnel à la fréquence de rencontre entre les prédateurs et les proies. Bien entendu, la prédation a tendance à faire croître la taille de la population de prédateurs (avec un taux $d$), et à faire décroître la taille de la population de proies (avec un taux $b$).

**Définition 1.2** (Trajectoire et orbite)
*La* trajectoire *d'un point $x$ de $S$, est l'application $F_x : T \to S$ qui associe à chaque temps $t$ de $T$ l'état $F_x(t) = F(x,t)$, qui est l'état global atteint après un temps $t$ à partir de l'état global initial $x$.*
*L'*orbite *d'un point $x$ de $S$ est définie comme l'ensemble : $O(x) = \{F(x,t),\ t \in T\}$.*

Dans la pratique, on confond trajectoire et orbite. Ainsi, une trajectoire désigne l'ensemble des état globaux atteints par le système à partir d'un état global initial donné. Lorsque l'on étudie un système dynamique, on s'intéresse aux différentes trajectoires que l'on peut obtenir à partir de différents états initiaux, et plus particulièrement au comportement asymptotique de ces trajectoires. Ces comportements asymptotiques représentent les états du système lorsque le temps s'écoule vers l'infini. Nous observons deux types de comportements asymptotiques : soit le système diverge, soit il adopte un comportement "stable", que l'on appelle attracteur. Le notion d'attracteur est une notion complexe en général. Des définitions complètes ont été données dans le cadre de système dynamiques réels [18, 19]. Dans le cadre de notre travail, nous n'avons pas besoin de définition si complexe et la définition simple suivante nous suffit. Nous supposons que $S$ est muni d'une distance $d$ équivalente à la distance euclidienne.

**Définition 1.3** (Attracteur et bassin d'attraction)
*$Attr \subseteq S$ est un* attracteur *pour le système dynamique $(S, T, F)$ si :*

1. *$Attr$ est non vide,*
2. *$Attr$ est invariant ($\forall t \in T,\ \ F(Attr, t) \subseteq Attr$),*
3. *il n'existe pas de sous-ensembe non vide $B \subset Attr$ qui vérifie 2.*

*Le* bassin d'attraction *associé à l'attracteur $Attr$, noté $B(Attr)$ est le sous-ensemble non vide de $S$ défini par : $B(Attr) = \{x \in S \mid L(x) \subset Attr\}$ avec $L(x) = \{y \in X \mid \forall \varepsilon > 0, \forall t_0 > 0, \exists t > t_0$ tel que $d(y, F(x,t)) < \varepsilon\}$.*

De façon informelle, le bassin d'attraction associé à un attracteur est l'ensemble des points dont la trajectoire "passe aussi près que possible" de l'attracteur. La principale différence entre notre définition et celles données dans [18, 19] est que nous considérons qu'un attracteur est inclus dans son bassin d'attraction, contrairement aux autres auteurs. Ainsi, selon notre définition, les comportements limites instables sont des attracteurs, dont le bassin d'attraction est réduit au seul attracteur. Cette différence dans la définition d'un attracteur

n'a en revanche pas d'incidence sur l'étude des systèmes de régulation tels que nous les modélisons.

La figure 1.1 illustre les notions de trajectoire, attracteur et bassin d'attraction en présentant trois trajectoires issues de différents états initiaux pour le système dynamique de l'exemple 1.1 (avec $a = 5$, $b = 3$, $c = 1$, $d = 1$). Les flèches indiquent le sens de parcours des trajectoires lorsque le temps s'écoule. Les trajectoires bleue et noire sont cycliques ; dans ce cas, trajectoire, attracteur et bassin d'attraction sont confondus. La trajectoire rouge correspond à un état initial pour lequel la taille de la population de proies est nulle. On voit que la taille de la population de prédateurs décroît alors jusqu'à devenir nulle également. Le point $(0, 0)$ est donc un attracteur pour ce système et le bassin d'attraction qui lui est associé est la droite $x = 0$. Il existe un autre attracteur, le point $(\frac{c}{d}, \frac{a}{b})$. On trouve ce point d'équilibre en posant $x'(t) = 0$ et $y'(t) = 0$. Dès que l'on s'écarte de ce point dans l'espace des états, on passe sur une trajectoire cyclique, donc le bassin d'attraction de cet attracteur est réduit à lui-même.

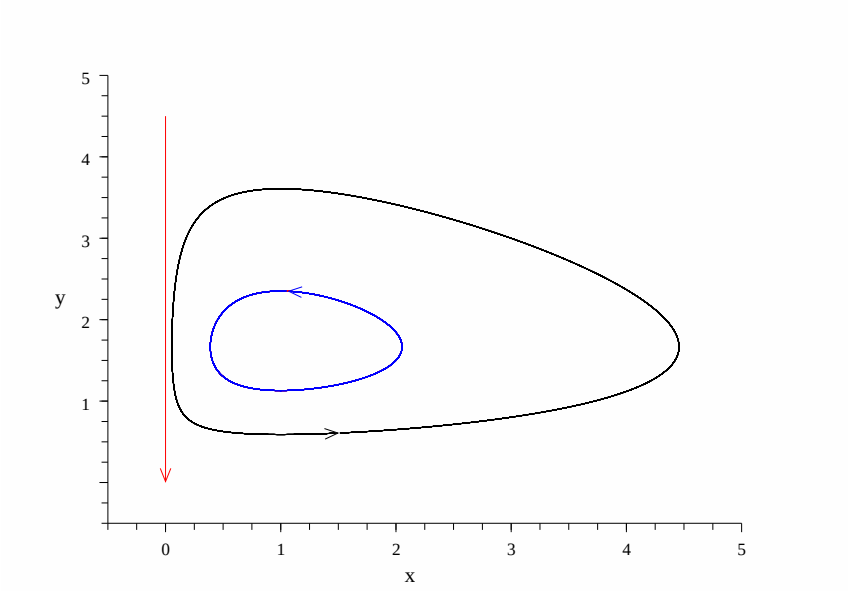



FIG. 1.1 – Trois trajectoires pour le système dynamique de l'exemple 1.1 avec les paramètres suivants : $a = 5$, $b = 3$, $c = 1$, $d = 1$

Nous avons restreint notre étude à une classe particulière de systèmes dynamiques, les systèmes dynamiques dont le temps évolue dans l'ensemble des entiers naturels. Ces systèmes dynamiques sont appelés systèmes dynamiques discrets (ou à temps discret).

**Définition 1.4** (Système dynamique discret)
*Un* système dynamique discret *est un système dynamique particulier* $(S, T, F)$ *où* $T = \mathbb{N}$.

Nous pouvons remarquer que le fait d'avoir réduit l'espace des temps à $\mathbb{N}$, nous permet d'écrire la fonction de transition globale $F$ comme l'itérée d'une fonction $G : S \to S$, c'est-à-dire : $\forall x \in S, \forall t \in \mathbb{N},\ \ F(x,t) = G^t(x)$.
En effet, en notant $\forall t \geq 1,\ F^t(x,1) = F(F^{t-1}(x,1),1)$ et $F^0(x,1) = x$, on a pour tout $t \geq 1$ :

$$\begin{aligned} F(x,t) &= F(x, 1+(t-1)) \\ &= F(F(x,1), t-1) \\ &= F(F^2(x,1), t-2) \\ &\vdots \\ &= F(F^{t-1}(x,1),1) \\ &= F^t(x,1). \end{aligned}$$

On a également $F(x,0) = x = F^0(x,1)$. Ainsi, $\forall x \in S, \forall t \in \mathbb{N},\ \ F(x,t) = F^t(x,1)$.

Pour la suite, nous considérons donc que la fonction de transition globale est une application de $S$ dans $S$, et non plus de $S \times \mathbb{N}$ dans $S$ (on remplace $F(x,t)$ par $F(x,1)$, que l'on note $F(x)$ par abus).

Remarquons enfin que, dans le cas où l'on modélise un phénomène réel par un système dynamique discret, on obtient une description qualitative du phénomène, alors qu'un modèle à base de système dynamique à temps continu aurait permis une description quantitative du même phénomène. En passant du continu au discret, on simplifie généralement le modèle, que l'on peut alors d'autant mieux étudier.

## 1.2 Réseaux d'automates et réseaux de régulation biologique

La modélisation discrète des réseaux de régulation biologique a été introduite dans les années 1940 par McCulloch et Pitts [20], avec leurs travaux sur les réseaux de neurones. Puis, à la fin des années 1960 et au début des années 1970, les recherches de Kauffman [21] et Thomas [22] mirent en avant l'intérêt de ces modèles pour la compréhension de la dynamique de réseaux de régulation génétique. La plupart des travaux ont focalisé sur un cas particulier de systèmes dynamiques discrets, les réseaux d'automates. Ceux-ci se sont avérés être de bons modèles pour les réseaux de régulation : ils permettent de simplifier au maximum les systèmes réels, de les modéliser avec une approche dynamique, tout en assurant un formalisme rigoureux. En effet, de façon informelle, en simplifiant à l'extrême, un élément d'un réseau de régulation peut être vu comme un objet dont l'état dépend d'un signal global qu'il reçoit (signal émis par les autres éléments du réseau). Les automates sont des objets mathématiques qui correspondent tout à fait à cette définition informelle.

### 1.2.1 Réseaux d'automates

**Définition 1.5** (Réseau d'automates)
*Un réseau d'automates $R$ composé de $n$ éléments est un couple $(Q, F)$ où :*
- *$Q \subset \mathbb{Z}$ est l'ensemble fini des états dans lesquels peut être chacun des $n$ automates,*
- *$F : Q^n \to Q^n$ est la fonction de transition globale*

*On note $F = (f_1, \cdots, f_n)$ et on appelle $f_i : Q^n \to Q$ la fonction de transition locale de l'élément $i$, qui permet la mise à jour de son état en fonction des états des éléments du réseau.*

On peut remarquer qu'à partir des éléments en interaction d'un réseau d'automates, on peut construire un graphe orienté.

**Définition 1.6** (Graphe d'interaction)
*On appelle graphe d'interaction associé au réseau d'automates $R = (Q, F)$ à $n$ éléments, le graphe orienté $G = (V, A)$ où :*
- *$V = \{1, \cdots, n\}$ est l'ensemble des sommets (ou noeuds) de $G$,*
- *$A \subseteq V \times V$ est l'ensemble des arcs de $G$, et est tel que :*

$$\forall i \in V,\ \forall j \in V,\ \ (i,j) \in A$$
$$\Updownarrow$$
$$\exists \mathbf{X} \in Q^n, y \in Q, y \neq x_i \ \ / \ f_j(x_1, \cdots, x_i, \cdots, x_n) \neq f_j(x_1, \cdots, y, \cdots, x_n)$$

Chacun des sommets du graphe d'interaction associé à un réseau d'automates représente un des automates du réseau et les arcs du graphe représentent les interactions entre éléments. Nous définissons le graphe d'interaction à partir des fonctions de transition locale. Ceci nous assure qu'un arc du graphe d'interaction entre le noeud $i$ et le noeud $j$ représente une réelle influence de l'automate $i$ sur l'automate $j$ dans le réseau. Ainsi, supprimer un arc du graphe d'interaction revient à modifier l'une des fonctions de transition locale du réseau et donc la dynamique du système. Formellement, on dit que le graphe d'interaction a une structure quasi-minimale. Nous reviendrons plus tard sur l'importance de cette notion de structure quasi-minimale.

**Notation 1.1** (Etat) **:** *Soit $R = (Q, F)$ un réseau d'automates composé de $n$ éléments en interaction.*
*On note $x_i(t)$ l'état de l'élément $i$ du réseau au temps $t$.*
*On note $\mathbf{X}(t) = (x_1(t), \cdots, x_n(t))$ l'état global du réseau au temps $t$. On appelle configuration du réseau d'automates tout état global du réseau.*

Il y a $(\#Q)^n$ configurations possibles pour un réseau d'automates à $n$ éléments.

*Exemple* 1.2 *:* La figure 1.2 présente le graphe d'interaction associé à un réseau de trois automates booléens (à états dans $Q = \{0, 1\}$), dont les fonctions de transition locale sont les fonctions booléennes suivantes (les opérateurs utilisés

sont les opérateurs booléens classiques) :

$$\begin{aligned} f_1(x_1, x_2, x_3) &= x_1 \wedge x_2 \\ f_2(x_1, x_2, x_3) &= x_1 \vee \neg x_3 \\ f_3(x_1, x_2, x_3) &= \neg x_1 \end{aligned}$$

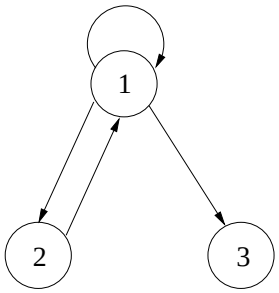



FIG. 1.2 – Graphe orienté associé au réseau d'automates booléens dont les fonctions de transition locale sont données dans l'exemple 1.2

On voit que la fonction de transition locale de l'élément 1, $f_1$ dépend des variables $x_1$ et $x_2$, mais pas de $x_3$. Dans le graphe d'interaction, le noeud 1 reçoit donc un arc provenant de lui-même et du noeud 2, mais pas du noeud 3. On peut faire le même raisonnement sur les fonctions $f_2$ et $f_3$ afin d'expliquer les arcs qui arrivent sur les noeuds 2 et 3 dans le graphe d'interaction.

Profitons d'avoir défini le graphe d'interaction associé à un réseau d'automates pour rappeler quelques notions relatives aux graphes orientés qui nous seront utiles. Pour les définitions qui suivent, nous considérons que $G = (V, A)$ est un graphe orienté.

**Définition 1.7** (Source et destination d'un arc)
*Soit $a = (i, j) \in A$ un arc de $G$.*
*Le sommet $i$ est appelé (sommet)* source *de $a$.*
*Le sommet $j$ est appelé (sommet)* destination *de $a$.*

**Définition 1.8** (Degré d'un noeud et connectivité d'un graphe)
*Soit $i \in V$ un sommet de $G$.*
*Le* degré entrant *de $i$ est le nombre d'arcs de $G$ dont $i$ est la destination.*
*Le* degré sortant *de $i$ est le nombre d'arcs de $G$ dont $i$ est la source.*
*La* connectivité *du graphe $G$ est définie comme étant le degré (entrant ou sortant) moyen de ses noeuds. Elle est égale à $\frac{\#A}{\#V}$.*

**Définition 1.9** (Noeud source et noeud puits)
*On dit d'un noeud d'un graphe orienté que c'est un* noeud source *si son degré entrant est nul.*
*On dit d'un noeud d'un graphe orienté que c'est un* noeud puits *si son degré sortant est nul.*

**Définition 1.10** (Chemin)
*Un sous-ensemble de $A$, noté $C = \{a_1, \cdots, a_r\}$ est appelé* chemin *si, pour tout $k \in \{2, \cdots, r\}$, la source de $a_k$ est la destination de $a_{k-1}$. Le chemin $C$ permet donc d'atteindre le noeud destination de $a_r$ à partir du noeud source de $a_1$.*

**Définition 1.11** (Composante fortement connexe)
*$\mathcal{C}$, un sous-ensemble de $V$, est appelé* composante fortement connexe *de $G$ si, pour chaque sommet de $\mathcal{C}$, il existe un chemin vers tous les autres sommets de $\mathcal{C}$, et $\mathcal{C}$ est de taille maximale, c'est-à-dire qu'il n'est pas strictement contenu dans un sous-ensemble possédant cette propriété.*
*On dit que $G$ est un* graphe fortement connexe *s'il a une unique composante fortement connexe qui contient tous les sommets de $G$.*

**Définition 1.12** (Circuit)
*Un* circuit *est un chemin $C = \{a_1, \cdots, a_r\}$ tel que la destination de $a_r$ est la source de $a_1$. On dit alors que $C$ est un circuit de longueur $r$.*
*Une* boucle *est un circuit de longueur* 1.

Le graphe de l'exemple 1.2, présenté sur la figure 1.2, contient une boucle et un circuit de longueur 2. La connectivité de ce graphe est égale à $\frac{4}{3}$ et ce n'est pas un graphe fortement connexe puisqu'il n'existe aucun chemin entre le noeud 3 et les autres sommets (le sous-ensemble de sommets $\{1, 2\}$ forme une composante fortement connexe, de même que le sous-ensemble réduit à un unique noeud $\{3\}$). Le noeud 3 est un noeud puits.

On peut définir quatre grandes familles de graphes parmi les graphes d'interaction les plus étudiés :

- les graphes aléatoires, proposés par Erdõs et Rényi [23, 24] ; la distribution des degrés des noeuds de ces graphes est caractéristique, c'est une loi de Poisson ;
- les graphes complets, dont tous les sommets sont connectés les uns aux autres ;
- les graphes cellulaires, dont les noeuds sont situés sur une grille à $d$ dimensions (en général, on a $d \leq 3$) ; chaque noeud est connecté à ses voisins sur la grille, et on a le même type de voisinage pour tous les noeuds ; ces graphes sont associés à un classe particulière de réseaux d'automates, les réseaux d'automates cellulaires [25, 26], pour lesquelles les fonctions de transition locales sont toutes identiques ;
- enfin, les graphes complexes, notamment étudiés par Albert et Barabási [27] ; dans le cadre des systèmes complexes, ce sont les graphes les plus étudiés ces dernières années, car ils correspondent aux graphes d'interaction de réseaux biologiques ou sociaux réels ; ces graphes ne sont pas aléatoires au sens d'Erdõs et Rényi ; la distribution des degrés des noeuds est ici encore caractéristique, c'est une loi puissance, c'est-à-dire qu'il existe un faible nombre de sommets à fort degré et un très grand nombre de sommets à faible degré.

D'un point de vue dynamique, la définition des attracteurs et bassins d'at-

traction est bien plus aisée pour les réseaux d'automates que pour les systèmes dynamiques en général. Nous supposons pour la suite que la fonction de transition globale d'un réseau d'automates $R = (Q, F)$ à $n$ éléments est une fonction déterministe, c'est-à-dire que pour chaque $\mathbf{X} \in Q^n$, il existe une unique image $F(\mathbf{X})$. Les travaux présentés dans ce manuscrit concernent uniquement les réseaux d'automates déterministes. Sené [25] propose une étude concernant les réseaux d'automates à fonction de transition globale stochastique.

Comme l'ensemble des états de chaque automate $Q$ est fini, il y a un nombre fini de configurations possibles pour le réseau. De plus, comme la fonction de transition globale est considérée conne étant déterministe, chaque trajectoire issue d'une configuration initiale est finie. A partir d'un certain nombre d'itérations de la fonction de transition globale, les configurations atteintes se répètent indéfiniment.

**Définition 1.13** (Attracteur et bassin d'attraction)
*Soit $R = (Q, F)$ un réseau d'automates à $n$ éléments.*
*Un sous-ensemble de $Q^n$, noté Attr est un* attracteur *de $R$ si les configurations de Attr se répètent indéfiniment par itération de la fonction de transition globale.*
*Le* bassin d'attraction *associé à l'attracteur Attr est l'ensemble des configurations contenues dans les trajectoires qui contiennent Attr.*

On voit que l'on peut définir deux types d'attracteurs pour les réseaux d'automates :

- les attracteurs composés d'au moins 2 configurations sont appelés des cycles limites ;
- les attracteurs qui contiennent 1 unique configuration sont appelés des points fixes.

Formellement, on définit les cycles limites et les points fixes comme suit :

**Définition 1.14** (Cycle limite)
*Soit $R = (Q, F)$ un réseau d'automates à $n$ éléments.*
*La suite de configurations $C \subseteq Q^n, C = [\mathbf{X}^1, \cdots, \mathbf{X}^l], l \geq 2$ est un* cycle limite *de période (ou longueur) $l$ pour $R$ si, et seulement si :*

- *les $\mathbf{X}^j$ sont deux à deux différents,*
- *$F(\mathbf{X}^j) = \mathbf{X}^{j+1}$ pour tout $j \in \{1, \cdots, l-1\}$ et*
- *$F(\mathbf{X}^l) = \mathbf{X}^1$.*

**Définition 1.15** (Point fixe)
*Soit $R = (Q, F)$ un réseau d'automates à $n$ éléments.*
*La configuration $\mathbf{X} \in Q^n$ est un* point fixe *pour $R$ si, et seulement si $F(\mathbf{X}) = \mathbf{X}$.*

Les points fixes d'un réseau d'automates sont les points fixes de sa fonction de transition globale.

### 1.2.2 Modèles de réseaux de régulation biologique

Nous avons vu que, dès le début des années 1970, Kauffman [21], puis Thomas [22] ont rendu légitime l'utilisation des réseaux d'automates comme modèles qua-

litatifs des réseaux d'interaction biologique, et plus particulièrement des réseaux de régulation génétique. Dans la suite du manuscrit, on évoquera la modélisation de réseaux de régulation génétique comme exemple de réseaux de régulation biologique. Kauffman et Thomas ont chacun proposé une modélisation sous forme de réseau d'automates. Chaque gène est modélisé par un automate dont l'état représente l'activité, l'expression du gène considéré. L'intérêt de ces modèles est que l'on peut donner une signification biologique aux attracteurs des réseaux d'automates. En effet, dans le cadre de la modélisation de réseaux de régulation génétique, depuis Kauffman [28], on associe les attracteurs du modèle avec les différents types cellulaires dans lesquels le réseau génétique est impliqué. On considère que chaque type de cellule différenciée est caractérisé par un certain phénotype (ensemble des caractères observables). Comme toutes les cellules d'un même organisme possèdent le même génotype (à part les cellules immunocompétentes, dont les gènes du répertoire immunologique sont différents, et à part les cellules pathologiques présenant des mutations ou des anomalies chromosomiques acquises), la différence de phénotype s'explique par une différence au niveau de l'activité protéique, enzymatique... La différence de phénotype s'explique donc par une différence d'expression des gènes. Ainsi, chaque type de cellule différenciée est caractérisé par un profil d'expression génétique, à l'équilibre. A chaque attracteur du réseau d'automates qui sert de modèle, correspond un profil d'états des automates du réseau. On fait le lien direct entre ce profil d'états et le profil d'activation des gènes. Les attracteurs d'un réseau d'automates modélisant un réseau de régulation génétique représentent tous un type cellulaire dans lequel le réseau de régulation est impliqué. Bien qu'il soit plus aisé de voir la correspondance entre un point fixe du réseau d'automates et un profil d'expression de gènes que l'on peut considérer qualitativement comme stable, un cycle limite peut également être associé directement à une fonction biologique, qui sera cyclique (voir l'exemple 1.4).

Nous précisons ici que ce manuscrit traite d'un certain type de modèle de réseau de régulation génétique, mais nous n'abordons pas le problème de reconstruction d'un tel réseau à partir des données biologiques. Cette étape de reconstruction (que l'on appelle également problème inverse) est une étape difficile qui nécessite, la plupart du temps, le recours à des méthodes d'optimisation (recuit simulé [29] ou algorithme génétique [30] par exemple). Nous n'entrerons pas dans de plus amples explications, et renvoyons à la lecture de [31, 32, 33, 34, 35] pour des détails supplémentaires.

### Modèles de Kauffman et de Thomas

Les deux types de réseaux d'automates introduits par Kauffman et Thomas, font partie, encore aujourd'hui, des modèles discrets les plus utilisés pour les réseaux de régulation génétique. Nous allons donc brièvement les décrire avant de présenter le modèle particulier sur lequel a porté l'étude présentée dans ce manuscrit. Pour une revue plus complète des différents modèles de réseaux de régulation génétique, consulter [13].

Afin de simplifier au maximum le modèle, Kauffman [21, 28] considère qu'un gène peut être modélisé par une variable booléenne qui représenterait l'activité du dit gène : le gène est actif (lorsque la variable vaut 1) ou inactif (lorsque la variable vaut 0), c'est-à-dire que les produits synthétisés à partir du gène sont présents ou absents. De plus, les interactions entre gènes peuvent être représentées par n'importe quelle fonction booléenne qui permet de calculer l'état du gène (actif ou inactif) à partir de l'activation des autres gènes du réseau. Enfin, on considère que la mise à jour des états de chaque automate se fait de manière synchrone. Le formalisme de modélisation proposé par Kauffman revient donc à utiliser un réseau d'automates booléens synchrones comme modèle de réseau d'interaction génétique. On parle plus communément de réseaux booléens synchrones.

Thomas [36, 37, 38] propose une modélisation sous forme de réseaux logiques que l'on appelle communément réseaux de Thomas. Chaque gène $i$ d'un réseau est représenté par une variable discrète, $\hat{x_i}$, appelée variable logique, qui peut prendre plusieurs valeurs. Le nombre de valeurs que la variable $\hat{x_i}$ peut prendre est lié au nombre d'éléments du système qui sont sous l'influence du gène $i$. A chaque élément $j$ sous l'influence du gène $i$ correspond une valeur pour la variable $\hat{x_i}$. Cette valeur représente le niveau d'activation minimal à partir duquel l'activation du gène $i$ a un effet sur l'élément $j$. Si le gène $i$ influence $k$ éléments, alors la variable $\hat{x_i}$ peut prendre jusqu'à $k+1$ valeurs possibles (0 compris). C'est une des différences principales qui existent avec les réseaux booléens dans lesquels il n'y qu'un niveau d'activation possible. L'autre différence majeure est, que dans les réseaux de Thomas, les états des automates sont mis à jour de façon asynchrone. Cela modélise le fait qu'il est très peu probable que les concentrations de produits synthétisés à partir de 2 gènes différents, passent chacune un de leurs seuils en même temps. Ainsi, à une configuration donnée d'un réseau de Thomas peut correspondre plusieurs configurations après itération de la fonction de transition globale puisqu'il y a autant de façons de réaliser la transition que d'éléments constituant le réseau. Concernant les attracteurs du réseau, cela rend l'étude des cycles limites délicate. En revanche, il n'y a aucun problème vis-à-vis des points fixes, puisque nous avons vu que tout point fixe du réseau correspond à un point fixe de la fonction de transition globale.

L'un des avantages des réseaux de Thomas sur les réseaux de Kauffman est que l'on se limite à des interactions entre 2 gènes qui sont simples. Dans le cas où le gène $i$ influence le gène $j$, cette influence peut être de 2 types différents : en présence de $i$, $j$ a tendance à être soit activé (on dira que $i$ est activateur de $j$), soit inhibé (on dira que $i$ est inhibiteur de $j$). Ces 2 types d'influence correspondent bien aux types d'interactions que l'on connaît entre gènes [39]. Dans le cas de réseaux booléens, on peut avoir des influences qui ne sont ni activation, ni inhibition : imaginons par exemple un réseau d'automates booléens dont l'automate $i$ est sous l'influence des automates $j$ et $k$ et tel que la fonction de transition locale correspondant à $i$ est $f_i = x_j \oplus x_k$[1] ; $j$ et $k$ ne sont ni activateurs, ni inhibiteurs de $i$. On peut définir une certaine classe de fonctions booléennes

[1] $\oplus$ désigne le ou exclusif booléen

qui permettent de ne travailler qu'avec des interactions de type activation ou inhibition.

**Définition 1.16** (Fonction booléenne monotone et définie par signe)
*Soit* $f : \{0,1\}^n \to \{0,1\}$ *une fonction booléenne.*
*On dit que* $f$ *est monotone croissante par rapport à sa* $i^e$ *variable si, et seulement si :*

$$\forall \mathbf{X} =\in \{0,1\}^n,\ f(x_1, \cdots, x_{i-1}, 0, x_{i+1}, \cdots, x_n) \leq f(x_1, \cdots, x_{i-1}, 1, x_{i+1}, \cdots, x_n).$$

*On dit que* $f$ *est monotone décroissante par rapport à sa* $i^e$ *variable si, et seulement si :*

$$\forall \mathbf{X} =\in \{0,1\}^n,\ f(x_1, \cdots, x_{i-1}, 0, x_{i+1}, \cdots, x_n) \geq f(x_1, \cdots, x_{i-1}, 1, x_{i+1}, \cdots, x_n).$$

*On dit que* $f$ *est définie par signe si elle est monotone croissante ou décroissante par rapport à chacune de ses variables.*

En ne travaillant qu'avec des réseaux d'automates booléens dont les fonctions de transition locale sont des fonctions définies par signe, on ne peut modéliser que des influences du type activation ou inhibition. En effet, si la fonction de transition locale de l'élément $i$, $f_i$, dépend de sa $j^e$ variable, et est monotone croissante (resp. décroissante) par rapport à cette variable, alors le gène modélisé par l'élément $j$ est activateur (resp. inhibiteur) du gène modélisé par l'élément $i$. On appellera par la suite réseau de régulation booléen tout réseau d'automates booléens dont les fonctions de transition locale sont des fonctions définies par signe. Ces dernières années, Kauffman a restreint son modèle à des réseaux de régulation booléens [40, 41]. Pour cela, il a introduit les fonctions canalisantes hiérarchiques [2], qui sont une sous-classe stricte de fonctions définies par signe. Nous n'entrerons pas ici dans la description de ces fonctions, mais nous noterons que l'utilisation de celles-ci permet donc de ne travailler que sur des réseaux dont les interactions sont de type activation ou inhibition.

### Réseaux d'automates booléens à seuil

Dans le cadre de notre étude, nous avons travaillé sur une classe particulière de réseaux d'automates booléens, les réseaux d'automates booléens à seuil. Les fonctions de transition locale de ces réseaux sont des fonctions booléennes monotones. Ces réseaux ont été introduits pour la première fois dans le contexte des réseaux de neurones formels en 1943 par McCulloch et Pitts [20]. Ils ont ensuite été rendus célèbres par les travaux de Hopfield [42, 43] sur la mémoire associative, au début des années 1980. On les retrouve utilisés plus récemment, dans le cadre de la modélisation des réseaux de régulation génétique, notamment concernant le réseau de régulation lié à la morphogénèse florale chez *Arabidopsis Thaliana* [44].

[2] Traduction littérale personnelle du terme anglais "hierarchical canalyzing functions"

**Définition 1.17** (Réseau d'automates booléens à seuil)
*Un réseau d'automates booléens à seuil (RABS), composé de $n$ éléments, est un réseau d'automates particulier $R = (G, F)$, tel que :*

- $G = \{0, 1\}$*, les automates sont des automates booléens*
- $F : \{0,1\}^n \to \{0,1\}^n$*, la fonction de transition globale est composée de $n$ fonctions à seuil ; la fonction de transition locale de l'élément $i$ est de la forme :*

$$\forall \mathbf{X} \in \{0,1\}^n, \quad f_i(\mathbf{X}) = H\left(\sum_{j=1}^{n} w_{ij}x_j - \theta_i\right),$$

*avec* $\forall j \in \{1, \cdots, n\}$, $w_{ij} \in \mathbb{Z}$ , $\theta_i \in \mathbb{Z}$, *et où $H$ est la fonction de Heaviside, définie sur $\mathbb{R}$ par* $H(x-\theta) = \begin{cases} 0 & si \quad x \leq \theta \\ 1 & si \quad x > \theta \end{cases}$.

Avant de donner une description et une interprétation des différents paramètres des fonctions de transition locale d'un RABS, nous allons nous intéresser au choix de la fonction de Heaviside.

Cette fonction, associée à un seuil $\theta$, peut être vue comme une fonction sigmoïdale dont le point d'inflexion serait situé au niveau du seuil, et dont la "pente" serait infinie. Rappelons qu'une fonction sigmoïdale est une fonction du type $f(x) = \frac{x^n}{1+x^n}$ ou du type $f(x) = \frac{1}{1+e^{-\lambda(x-\theta)}}$. On retrouve les fonctions sigmoïdales dans le domaine de la cinétique enzymatique : sous certaines conditions, la vitesse d'une réaction catalysée par une enzyme en fonction de la concentration en substrat (de l'enzyme) est d'allure sigmoïdale. La figure 1.3 qui présente le graphe d'une fonction de Heaviside (en noir) associée à un seuil $\theta$, ainsi que les graphes de deux fonctions sigmoïdales (en bleu et en rouge) dont les points d'inflexion sont situés au niveau du même seuil $\theta$, montre que l'on peut considérer la fonction de Heaviside comme l'équivalent d'une fonction sigmoïdale booléenne. Cette remarque justifie l'utilisation de la fonction de Heaviside dans la modélisation de la dynamique des réseaux de régulation génétique.

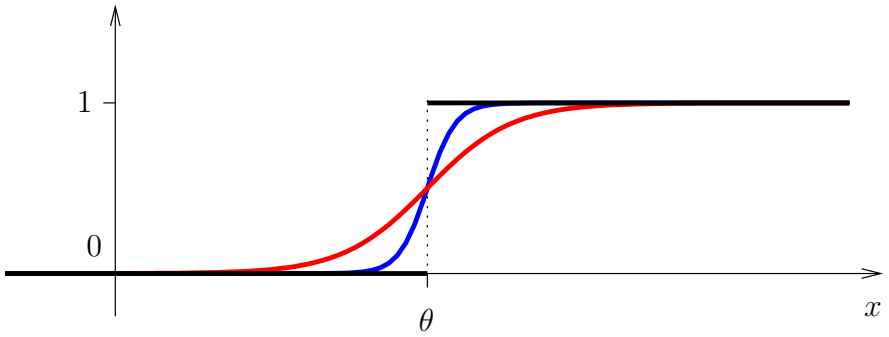



FIG. 1.3 – Graphes de la fonction de Heaviside (en noir) et de deux fonctions sigmoïdales ($\lambda = 2$ en rouge et $\lambda = 6$ en bleu).

Toujours concernant la fonction de Heaviside, notons que certains auteurs

([25] par exemple) utilisent une autre définition de la fonction de Heaviside, que l'on note $H^-$ : $H^-(x-\theta) = \begin{cases} 0 & \text{si} \quad x < \theta \\ 1 & \text{si} \quad x \geq \theta \end{cases}$. La seule différence entre $H$ et $H^-$ est la valeur de l'image de $\theta$. Comme nous travaillons avec des arguments entiers, on a $\forall (x, \theta) \in \mathbb{Z}^2$,
$H(x-\theta) = H^-(x-\theta+1)$. Les RABS définis à partir des 2 fonctions de Heaviside différentes sont donc équivalents, il suffit d'ajuster les seuils afin de passer de l'un à l'autre.

Enfin, d'autres auteurs ([45] par exemple) utilisent un type de réseaux d'automates, que l'on appelle réseaux d'automates à signe, pour modéliser les réseaux de régulation génétique. Ces réseaux sont très proches des RABS tels que nous les avons définis. Dans un réseau d'automates à signe, les automates sont à états dans $\{-1, 1\}$ (ce ne sont donc pas des automates booléens) et les fonctions de transition locales sont définies comme suit :

$$\forall \mathbf{Y} \in \{-1,1\}^n, \quad g_i(\mathbf{Y}) = sign\left(\sum_{j=1}^{n} \tilde{w}_{ij} y_j - \tilde{\theta}_i\right),$$

avec $\tilde{w}_{ij} \in \mathbb{Z}$, $\tilde{\theta} \in \mathbb{Z}$, et où $sign$ est la fonction signe, définie sur $\mathbb{R}$ par
$sign(x-\theta) = \begin{cases} -1 & \text{si} \quad x \leq \theta \\ 1 & \text{si} \quad x > \theta \end{cases}$.
On peut montrer que tout réseau d'automates à signe est équivalent à un réseau d'automates booléens à seuil. En effet, en posant le changement de variables suivant :

$$\begin{aligned} \mathbf{Y} &= 2\mathbf{X} - \mathbf{1}, \\ \tilde{w}_{ij} &= w_{ij}, \\ \tilde{\theta}_i &= 2\theta_i - \sum_{j=1}^{n} w_{ij}, \end{aligned}$$

on a :

$$\begin{aligned} \sum_{j=1}^{n} \tilde{w}_{ij} y_j - \tilde{\theta}_i &= \sum_{j=1}^{n} w_{ij}(2x_j - 1) - (2\theta_i - \sum_{j=1}^{n} w_{ij}) \\ &= 2\sum_{j=1}^{n} w_{ij} x_j - 2\theta_i - \sum_{j=1}^{n} w_{ij} + \sum_{j=1}^{n} w_{ij} \\ &= 2\left(\sum_{j=1}^{n} w_{ij} x_j - \theta_i\right). \end{aligned}$$

Comme $\forall x \in \mathbb{R},\ H(x) = H(2x)$, on a une égalité entre les fonctions de transition locale ($\forall \mathbf{Y} \in \{-1,1\}^n,\ g_i(\mathbf{Y}) = 2f_i(\frac{\mathbf{Y}+\mathbf{1}}{2}) - 1$) et donc ,en associant les états $-1$ et 0, les réseaux d'automates ont exactement le même comportement.

Nous avons montré que le choix précis de la fonction de Heaviside n'était pas important, puisque l'on pouvait toujours se ramener au même réseau d'automates. Voyons maintenant ce que représentent les divers paramètres utilisés dans la définition des fonctions de transition locale.

Dans un RABS, les $w_{ij}$ sont appelés les poids d'interaction. Le poids d'interaction $w_{ij}$ représente l'influence de l'élément $j$ sur l'élément $i$. Il est non nul si (l'état de) l'élément $i$ dépend de (l'état de) l'élément $j$. Si le RABS sert de modèle pour un réseau de régulation génétique, alors, $w_{ij}$ est positif (resp. négatif) si le gène $j$ est activateur (resp. inhibiteur) du gène $i$. L'amplitude du poids $w_{ij}$, sa valeur absolue $|w_{ij}|$, dépend de l'importance relative du gène $j$ dans la régulation du gène $i$. On détermine donc les amplitudes des poids, $|w_{ij}|$, en fixant tout d'abord $i$, puis en étudiant l'importance relative des différents éléments du réseau qui agissent dans la régulation du gène $i$ : plus forte est l'importance relative de l'élément $j$, plus grande sera $|w_{ij}|$. Il est important de noter que comparer les amplitudes des poids $w_{ij}$, pour une valeur de $j$ donnée, n'a pas de sens. Nous reviendrons sur cette remarque dans la partie 3.3.

Les $\theta_i$ sont appelés seuils d'activation. Le seuil $\theta_i$, associé à l'élément $i$ du réseau, représente la quantité totale d'activation à dépasser pour rendre l'élément $i$ actif : plus $\theta_i$ est élevé, plus l'élément $i$ est difficile à activer. Le signe du seuil d'activation $\theta_i$ indique l'état d'activation de l'élément $i$ lorsqu'il n'est soumis à aucune influence : $\theta_i < 0$ (resp. $\theta_i \geq 0$) signifie que l'élément $i$ est actif (resp. inactif), lorsqu'il n'est sous l'influence d'aucun élément du réseau.

Notons que nous ne considérons que des réseaux dont les poids d'interaction et les seuils sont entiers. Cette restriction n'a aucun effet sur le nombre de RABS que l'on peut ainsi définir. A chaque fonction booléenne à seuil définie par des poids d'interaction et un seuil réels, on peut faire correspondre des poids et un seuil entiers. Nous aborderons ce point dans la partie 2.2.

Dans un RABS, l'activation du sommet $i$, décrite par la fonction de transition locale $f_i$, se fait si la "somme des activations et inhibitions" qu'il reçoit de la part du réseau est supérieure à son seuil d'activation. Le poids d'interaction représentant l'influence du gène $j$ sur le gène $i$ est constant. Cela signifie donc que l'importance relative du gène $j$ dans la régulation du gène $i$ est la même quels que soient les autres gènes régulateurs de $i$ qui sont activés. Dans ce modèle, on peut donc ordonner les gènes activateurs (resp. inhibiteurs) d'un même gène $i$ par ordre d'importance dans l'activation (resp. l'inhibition) de $i$, en fonction de l'ordre sur les poids d'interaction. Supposons par exemple, que les gènes $j$ et $k$ soient activateurs du gène $i$. S'il existe un profil d'expression des autres gènes activateurs/inhibiteurs du gène $i$, pour lequel, lorsque $j$ est actif et $k$ inactif, il y a activation de $i$, et lorsque $j$ est inactif et $k$ actif, il n'y a pas activation de $i$, alors, on en déduit que $j$ est un activateur de $i$ plus important que $k$. Ainsi, il ne peut exister d'autre profil d'expression des autres gènes activateurs/inhibiteurs du gène $i$ pour lequel lorsque $j$ est inactif et $k$ actif, il y a activation de $i$, et lorsque $j$ est actif et $k$ inactif, il n'y a pas activation de $i$. Formellement, toute fonction booléenne à seuil est une fonction booléenne définie par signe, mais l'inverse n'est

pas vrai. L'exemple suivant présente une fonction définie par signe qui n'est pas une fonction booléenne à seuil.

*Exemple* 1.3 *:* On considère la fonction booléenne à 4 variables suivante : $f(\mathbf{X}) = (x_1 \wedge x_2) \vee (x_3 \wedge x_4)$. Il est facile de vérifier que $f$ est monotone croissante par rapport à chacune de ses variables. C'est donc une fonction définie par signe. Supposons que l'on puisse exprimer $f$ sous la forme d'une fonction à seuil, $f(\mathbf{X}) = H(w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 - \theta)$.
On a $f(0,0,1,1) = 1$ et $f(1,0,1,0) = 0$, d'où $\begin{cases} w_3 + w_4 > \theta \\ w_1 + w_3 \leq \theta \end{cases}$ et donc $w_1 < w_4$.
De même, de $f(1,1,0,0) = 1$ et $f(0,1,0,1) = 0$, on déduit $w_4 < w_1$. Il y a contradiction, $f$ n'est donc pas une fonction à seuil.

On remarque que chaque fonction de transition locale d'un RABS à $n$ éléments est entièrement définie par la donnée de $n$ poids d'interaction et d'un seuil d'activation. Ainsi, pour la suite du manuscrit, tout RABS sera naturellement défini par une matrice de dimensions $n \times n$, dite matrice d'interaction $\mathbf{W}$, dont le terme général est $[\mathbf{W}]_{ij} = w_{ij}$ et par un vecteur de dimension $n$, dit vecteur seuil, $\boldsymbol{\theta}$ dont le terme général est $[\boldsymbol{\theta}]_i = \theta_i$. De plus, le graphe d'interaction d'un RABS peut-être directement déterminé à partir de la matrice d'interaction (aussi appelée matrice d'incidence ou matrice adjacente) associée à ce réseau. Ceci nous permet de donner, pour les RABS, une définition du graphe d'interaction plus complète que la définition précédente (1.6), en ajoutant un signe à chaque arc du graphe.

**Définition 1.18** (Graphe d'interaction d'un réseau d'automates booléens à seuil) *Le graphe d'interaction associé au réseau d'automates booléens à seuil à n éléments défini par la matrice d'interaction* $\mathbf{W} = (w_{ij})$ *et le vecteur seuil* $\boldsymbol{\theta}$*, est le graphe orienté et signé* $G = (V, A)$ *tel que :*

- $V = \{1, \cdots, n\}$ *est l'ensemble des noeuds de G,*
- $A \subseteq V \times V \times \{-, +\}$ *est l'ensemble des arcs signés de G, et est tel que :*

$$\forall i \in V,\ \forall j \in V \quad \begin{cases} w_{ij} > 0 \Leftrightarrow (j, i, +) \in A \text{ (l'arc est positif)} \\ w_{ij} < 0 \Leftrightarrow (j, i, -) \in A \text{ (l'arc est négatif)} \end{cases}$$

De façon informelle, pour déterminer les arcs qui pointent sur le noeud $i$ du graphe d'interaction d'un réseau d'automates booléens à seuil, on parcourt la ligne $i$ de la matrice d'interaction de ce réseau. Chaque coefficient non nul $w_{ij}$ de cette ligne indique qu'il y a un arc entre le noeud $j$ et le noeud $i$. Le signe de l'arc est défini à partir du signe du coefficient de la matrice.

*Exemple* 1.4 *:* Chez le poulet, la morphogénèse de la plume est contrôlée par un réseau de régulation génétique que l'on peut réduire à 4 gènes principaux : *Wnt2*, *BMP2*, *BMP7* et la *Follistatine* (ou plutôt le gène qui code pour la protéine *Follistatine*) [46, 47]. Le graphe d'interaction de la figure 1.4 décrit les interactions entre ces gènes. Nous utilisons la convention suivante pour représenter les arcs

d'un graphe orienté et signé : les arcs positifs sont représentés par des flèches à bout noir, les arcs négatifs, par des flèches à bout blanc. La formation des primordia (ou "spots") précurseurs des plumes sur le dos de l'embryon de poulet (visibles sur la figure 1.4) est régie par l'expression des 3 gènes du réseau, *BMP2*, *BMP7* et la *Follistatine*. La présence des produits synthétisés à partir de ces gènes sous forme de spots, et non pas de manière uniforme, s'explique par une expression cyclique de ces gènes, associée à une diffusion des produits synthétisés. Si l'expression des gènes était constante, on ne pourrait expliquer la formation des spots. Ainsi, le réseau de régulation admet un comportement cyclique. Ce comportement cyclique se retrouve dans le modèle sous forme de RABS décrit par la matrice d'interaction et le vecteur seuil de la figure 1.4 (les gènes sont ordonnés de la manière suivante : *Wnt2*, *BMP7*, *BMP2*, *Follistatine*). En effet, ce réseau d'automates admet uniquement un cycle limite et pas de point fixe. Dans ce modèle, chaque activation (resp. inhibition) est modélisée par un coefficient 1 (resp. -1) dans la matrice d'interaction. Les seuils sont choisis nuls, sauf pour *Wnt2*, pour lequel on prend un seuil égal à -1 afin de fixer l'expression de *Wnt2* à 1 (le gène est actif).

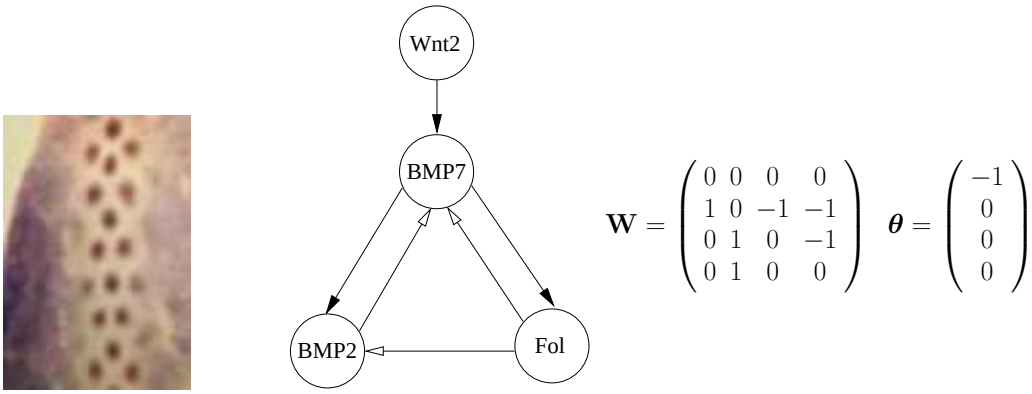



FIG. 1.4 – A gauche, spots précurseurs de plumes sur le dos d'un embryon de poulet ; au centre, graphe d'interaction du réseau de régulation génétique impliqué dans la morphogénèse de la plume chez le poulet ; à droite, matrice d'interaction et vecteur seuil du réseau d'automates booléens à seuil modèle de ce réseau de régulation (pour la matrice d'interaction et le vecteur seuil, les gènes sont ordonnés comme suit : *Wnt2*, *BMP7*, *BMP2*, *Follistatine*).

Du fait que l'on puisse définir tout RABS par une matrice d'interaction et un vecteur seuil, dont on extrait directement un graphe d'interaction orienté et signé, ces réseaux s'avèrent être de très bons réseaux d'étude que l'on peut facilement générer, dans le cadre de la modélisation des réseaux de régulation génétique.

Comme nous le faisons pour les RABS, nous pouvons signer le graphe d'interaction de tout réseau d'automates pour lesquels les interactions entre 2 éléments sont soit une activation, soit une inhibition. D'après ce que nous avons écrit précé-

demment, on peut donc signer le graphe d'interaction de tout réseau de Thomas, ainsi que les graphes des réseaux d'automates booléens si l'on se restreint aux réseaux de régulation booléens.

Dans un graphe signé, un chemin constitué d'un nombre pair (resp. impair) d'arcs négatifs est dit positif (resp. négatif). De même pour les circuits et les boucles.

### Conjectures et résultats importants

Il existe un grand nombre de résultats et conjectures connus concernant les graphes d'interaction de réseaux de régulation et les attracteurs de ces réseaux. Nous évoquons ici ceux qui nous semblent les plus importants, en commençant par des résultats sur les graphes d'interaction.

Tout d'abord, nous avons déjà vu dans la partie 1.2.1 que les graphes d'interaction des réseaux biologiques faisaient partie de la famille des graphes complexes. Les réseaux de régulation génétique sont également caractérisés par 2 statistiques concernant leur graphe d'interaction [45, 48] : la connectivité, ainsi que l'indice moyen d'inhibition (rapport du nombre d'arcs négatifs sur le nombre total d'arcs). En effet, pour beaucoup de réseaux de régulation génétique connus, la connectivité est comprise entre $\frac{3}{2}$ et 3 et l'index moyen d'inhibition est compris entre $\frac{1}{3}$ et $\frac{2}{3}$ ; c'est par exemple le cas pour l'opéron lactose, l'opéron *Cro* pour le phage $\lambda$, ou encore le réseau de la morphogénèse florale chez *Arabidopsis Thaliana*.

De récents travaux [49, 50] ont également montré que les graphes des réseaux de régulation génétique avaient une structure particulière. Ces réseaux comportent certains motifs, ou sous-réseaux de petite taille (de l'ordre de 3 ou 4 gènes) récurrents. Ces motifs sont présents dans les réseaux réels avec une fréquence bien plus élevée que dans des réseaux générés aléatoirement. Chacun de ces sous-réseaux a une fonction précise dans le processus d'expression génétique, et on peut les voir comme des modules élémentaires à assembler pour former un réseau complet. Par exemple, l'étude du réseau des interactions transcriptionnelles directes chez *Escherichia coli* a mis en évidence le fait que le réseau était composé de répétitions de 3 motifs particuliers [51]. Cette structure particulière des réseaux de régulation montre l'intérêt des travaux sur de petits réseaux d'étude, même si les réseaux réels sont des réseaux de (très) grande taille. La figure 1.5 présente le graphe d'interaction de 2 motifs : les "feed-forward loops" cohérente et incohérente de type 1, le type le plus représenté dans les réseaux transcriptionels chez *Escherichia coli* et *Saccharomyces cerevisiae* [52].

A partir de ses premiers travaux expérimentaux sur les réseaux booléens aléatoires [21], Kauffman a émis une conjecture [28] établissant un lien entre le nombre d'attracteurs d'un réseau de régulation génétique et le nombre de noeuds de ce réseau : le nombre d'attracteurs serait de l'ordre de la racine carrée du nombre de noeuds. Si l'on se rappelle de l'interprétation des attracteurs d'un réseau de régulation comme types cellulaires, on remarque que cette conjecture est tout à fait valide chez l'être humain. En effet, le génome humain est composé d'environ 35 000 gènes et on observe chez l'homme environ 200 tissus cellulaires différents.




FIG. 1.5 – "Feed forward loop" cohérente de type 1 (à gauche) et incohérente de type 1 (à droite)

Or, la racine carrée de 35 000 est environ égale à 187. La conjecture de Kauffman a été démontrée dans le cas de réseaux dont la connectivité est égale à 2 et en se restreignant aux seuls points fixes [45, 53]. En prenant en compte les cycles limites, cette conjecture s'est avérée être fausse puisque le nombre de cycles limites peut être beaucoup plus grand [54, 55, 56, 57] : la croissance du nombre de cycles limites serait plus rapide que linéaire.

Deux autres conjectures concernant les réseaux de régulation ont été émises par Thomas [58]. Ces deux conjectures lient le graphe d'interaction d'un réseau de régulation et les propriétés dynamiques de ce réseau. La première conjecture établit que la présence d'un circuit positif dans le graphe d'interaction est une condition nécessaire à la multi-stationnarité (existence de plusieurs points fixes). Plusieurs preuves de cette conjecture ont été données pour des systèmes continus [59, 60, 61, 62, 63]. La deuxième conjecture établit que la présence d'un circuit négatif dans le graphe d'interaction est, quant à elle, une condition nécessaire à l'existence d'oscillation stables (cycles limites). Cette conjecture a, elle aussi, été démontrée pour des systèmes continus [59, 60, 61, 64]. Dans le cadre des réseaux d'automates, les deux conjectures ont également été récemment prouvées, en se restreignant aux RABS tout d'abord [45, 53], puis en généralisant aux réseaux d'automates booléens [65], et enfin aux systèmes dynamiques discrets [66].

Enfin, nous pouvons citer les travaux [67, 68] qui donnent un majorant du nombre de points fixes pour un réseau. Dans [67], les réseaux sont du type "AND-OR" (chaque fonction de transition locale peut s'écrire soit comme une conjonction, soit comme une disjonction de ses variables) et le majorant est donné en fonction du nombre d'éléments du réseau. Dans [68], les réseaux sont des réseaux de régulation booléens, et le majorant du nombre de points fixes est donné pour tout réseau dont le graphe d'interaction contient au moins un circuit positif. Ce majorant est exprimé en fonction du nombre de circuits positifs du graphe.

## 1.3 Modes d'itération et robustesse

Dans la section précédente, nous avons défini ce qu'était un réseau d'automates et nous avons montré qu'un réseau d'automates booléens à seuil pouvait être un bon modèle pour la dynamique des réseaux de régulation génétique. Nous avons abordé le lien qui pouvait exister entre le graphe d'interaction d'un réseau

d'automates et les attracteurs de celui-ci. Dans cette section, nous allons présenter plus précisément les propriétés dynamiques des RABS, et voir la conséquence de ces propriétés en terme de robustesse pour ces réseaux.

### 1.3.1 Modes d'itération

En passant d'un formalisme à temps continu pour les systèmes dynamiques réels à un formalisme à temps discret pour les systèmes dynamiques discrets, on découple en quelque sorte le temps et l'espace des états. Ainsi, pour les systèmes comportant plusieurs variables, comme les réseaux d'automates par exemple, la mise à jour de l'état de chaque variable peut se faire de façon synchrone, c'est-à-dire que tous les états sont mis à jour en même temps, ou non.

Nous avons vu en 1.2.2 que, dans le formalisme de Kauffman, la mise à jour des états se fait de façon synchrone. Cela signifierait qu'il existe une sorte d'horloge interne à chaque organisme sur laquelle serait synchronisé tout le système d'expression génétique. Cette hypothèse ne paraît pas vraiment réaliste. C'est la raison pour laquelle, dans le formalisme de Thomas, la mise à jour des états se fait de façon asynchrone, avec un seul automate qui change d'état à chaque pas de temps. Pour les RABS, nous travaillons avec une classe de modes d'itération généraux, que l'on appelle modes d'itération blocs-séquentiels (ou encore modes d'itération série-parallèle). Dans ce type de mode d'itération, les éléments du réseau sont regroupés en blocs ordonnés (chaque élément fait partie d'un unique bloc), de telle sorte que les éléments d'un même bloc sont mis à jour de façon synchrone, alors que les blocs sont traités de façon séquentielle. A chaque pas de temps, tous les blocs sont traités, et donc l'état de tous les éléments est mis à jour. Ces modes d'itération blocs-séquentiels sont applicables à tout réseau d'automates, et pas uniquement aux RABS. Formellement, on définit les modes d'itération blocs-séquentiels comme suit.

**Définition 1.19** (Préordre total)
*Soit $\preceq$ une relation binaire définie sur un ensemble $E$.*
*$\preceq$ est un* préordre *si elle est*
- *réflexive, c'est-à-dire* $\forall a \in E,\ a \preceq a$,
- *et transitive, c'est-à-dire* $\forall (a,b,c) \in E^3,\ \left\{ \begin{array}{l} a \preceq b \\ b \preceq c \end{array} \right. \Rightarrow a \preceq c.$

*$\preceq$ est un ordre si c'est un* préordre *antisymétrique, c'est-à-dire*
$\forall (a,b) \in E^2, \left\{ \begin{array}{l} a \preceq b \\ b \preceq a \end{array} \right. \Rightarrow a = b.$
*$\preceq$ est un* préordre total *si c'est un préordre et que tous les éléments de $E$ sont comparables, c'est-à-dire* $\forall (a,b) \in E^2,\ a \preceq b$ *ou* $b \preceq a$.

On dit que deux éléments $a$ et $b$ d'un ensemble $E$ sont égaux selon le préordre $\preceq$ défini sur $E$, si on a $a \preceq b$ et $b \preceq a$. Au contraire, on dit que $a$ et $b$ sont strictement différents selon $\preceq$ si on a soit $a \preceq b$, soit $b \preceq a$, mais pas les deux en même temps (ou exclusif).

**Définition 1.20** (Modes d'itération blocs-séquentiels)

*Soit $R = (Q, F)$ un réseau d'automates à $n$ éléments, et $\preceq$ un préordre total sur l'ensemble $\{1, \cdots, n\}$.*
*On note $f_i$ la fonction de transition locale de l'élément $i$.*
*On dit que $R$ est itéré avec le* mode d'itération blocs-séquentiel *associé à $\preceq$ quand on a :*

$$\forall i \in \{1, \cdots, n\},\ \forall t \in \mathbb{N}, \quad x_i(t+1) = f_i\left(x_1(r_1), \cdots, x_n(r_n)\right)$$

$$\text{avec } r_j = \begin{cases} t & \text{si } i \preceq j \\ t+1 & \text{sinon} \end{cases}$$

Deux cas particuliers de modes d'itération blocs-séquentiels découlent directement de cette définition.

**Définition 1.21** (Mode parallèle et modes séquentiels)
*Si le préordre $\preceq$ est tel que tous les éléments de $\{1, \cdots, n\}$ sont égaux selon $\preceq$, alors tous les éléments sont mis à jour de façon synchrone. On appelle ce mode d'itération, le mode parallèle. En mode parallèle, on a :*

$$\forall i \in \{1, \cdots, n\},\ \forall t \in \mathbb{N}, \quad x_i(t+1) = f_i\left(x_1(t), \cdots, x_n(t)\right).$$

*Si, au contraire, les éléments de $\{1, \cdots, n\}$ sont 2 à 2 strictement différents selon $\preceq$, c'est à dire si $\preceq$ est un ordre, alors les éléments sont mis à jour de façon séquentielle, en tenant compte de l'ordre défini par $\preceq$. On appelle ces modes d'itération, les modes séquentiels. En supposant que les éléments de $\{1, \cdots, n\}$ sont rangés selon $\preceq$, en mode séquentiel, on a :*

$$\begin{array}{rll} & \forall t \in \mathbb{N}, & x_1(t+1) = f_1\left(x_1(t), \cdots, x_n(t)\right) \\ \forall i \in \{2, \cdots, n\}, & \forall t \in \mathbb{N}, & x_i(t+1) = f_i\left(x_1(t+1), \cdots, x_{i-1}(t+1), x_i(t), \cdots, x_n(t)\right) \end{array}$$

On peut remarquer que chaque préordre total défini sur l'ensemble $\{1, \cdots, n\}$ est associé à une unique partition ordonnée de $\{1, \cdots, n\}$. On peut donc également définir les modes d'itération blocs-séquentiels à partir des partitions ordonnées, comme dans [25] par exemple. Le mode parallèle correspond au cas où la partition est réduite à l'ensemble $\{1, \cdots, n\}$, alors que les modes séquentiels correspondent aux partitions ordonnées à $n$ éléments (chaque élément de la partition contient un unique élément de $\{1, \cdots, n\}$). Dans la suite du manuscrit, on désignera un mode d'itération blocs-séquentiel par la partition ordonnée à laquelle il est associé. Par exemple, pour les réseaux d'automates à 3 éléments, le mode d'itération (1)(3)(2) est le mode séquentiel pour lequel l'élément 1 est mis à jour en premier, puis l'élément 3, et enfin l'élément 2. Remarquons à nouveau que, malgré l'emploi de la notion d'ordre de mise à jour, les éléments sont tous mis à jour à chaque pas de temps, même pour les modes séquentiels.

A partir de tout préordre total défini sur $\{1, \cdots, n\}$ (ou de toute partition ordonnée de $\{1, \cdots, n\}$), on peut définir un mode d'itération blocs-séquentiel pour les réseaux d'automate à $n$ éléments. Le nombre de modes d'itération blocs-séquentiels pour les réseaux d'automates à $n$ éléments, noté $B(n)$, est donc égal

au nombre de préordres totaux que l'on peut définir sur $\{1, \cdots, n\}$ (ou au nombre de partitions ordonnées de $\{1, \cdots, n\}$). Ce nombre vérifie la relation de récurrence suivante :

$$B(n) = \sum_{i=0}^{n-1} \binom{n}{i} B(i), \text{ avec } B(0) = 1 \text{ et } \binom{n}{i} \text{ coefficient binomial.}$$

Les nombres $B(n)$ sont connus sous le nom de nombres de Fubini ou nombres de Bell avec ordre[3] [69]. Quelques uns de ces nombres sont donnés dans le tableau 1.1 pour de petites valeurs du nombre d'éléments $n$.

TAB. 1.1 – Nombre de modes d'itérations blocs-séquentiels pour des réseaux d'automates à $n$ éléments pour quelques petites valeurs de $n$.

| $n$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 10 | 12 |
|---|---|---|---|---|---|---|---|---|---|
| $B(n)$ | 1 | 3 | 13 | 75 | 541 | 4683 | 47293 | $1,02 \times 10^8$ | $2,81 \times 10^{10}$ |

On voit que le nombre de modes d'itération blocs-séquentiels pour des réseaux d'automates à $n$ éléments croît très rapidement avec $n$. Dans [70], Wilf montre qu'asymptotiquement, on a :

$$B(n) = \frac{1}{2(\ln 2)^{n+1}} n! + O((0,16)^n n!)$$

On a donc une croissance exponentielle du nombre de modes d'itération blocs-séquentiels pour les réseaux d'automates en fonction du nombre d'éléments de ces réseaux. L'approximation donnée par Wilf est valable même pour de petites valeurs de $n$ où elle reste très bonne (par exemple pour $n = 3$, $\frac{1}{2(\ln 2)^{n+1}} n! = 12,996$).

On remarquera enfin qu'il y a un unique mode parallèle quelle que soit la taille du réseau, mais que le nombre de modes séquentiels pour un réseau à $n$ éléments est $n!$ (nombre de permutations des $n$ éléments).

Biologiquement, on peut interpréter les modes d'itération blocs-séquentiels en terme de régulation génétique. Dans la phase de transcription, la chromatine, qui est le support des gènes dans le noyau, est remodelée pour permettre l'action de l'ARN polymérase, enzyme qui catalyse la réaction de transcription. Deux gènes $a$ et $b$ qui se retrouvent proches lors de ce remodelage de la chromatine sont susceptibles d'être transcrits au même moment. Ainsi, au niveau de son expression, le gène $b$ ne peut subir l'effet de l'ARN messager qui est en train d'être produit à partir du gène $a$ et encore moins des protéines obtenues après traduction de cet ARN messager. L'expression du gène $b$ dépendra donc de l'état d'activation antérieur de $a$. Du point de vue du modèle, on peut donc considérer que les gènes $a$ et $b$ sont dans un même bloc du mode d'itération blocs-séquentiel.

[3] Traduction littérale personnelle du terme anglais "ordered Bell numbers"

L'ordre des blocs peut ensuite être déterminé en fonction de l'ordre dans lequel on observe que l'expression des gènes a lieu, comme dans [44].

### 1.3.2 Robustesse

Dans la partie précédente, nous avons montré que l'on pouvait itérer un réseau d'automates de plusieurs façons possibles. Les attracteurs d'un réseau, cycles limites et points fixes, sont plus ou moins sensibles à des changements de mode d'itération. Nous avons déjà vu que les points fixes de la dynamique d'un réseau d'automates correspondent à des points fixes de sa fonction de transition globale. On voit donc que les points fixes sont toujours les mêmes, quel que soit le mode d'itération choisi [71, 72]. En revanche, les cycles limites peuvent être sensibles à des changements de mode d'itération. Tant leur nombre, que leur nature même, peut varier avec les modes d'itération : pour un même réseau d'automates, on peut avoir des cycles limites pour un certain type de modes d'itération, et n'avoir que des points fixes pour un autre type de modes d'itération. Les cycles limites observés pour deux modes d'itération différents et pour un même réseau d'automates ne sont pas non plus toujours les mêmes. Ainsi, une perturbation dans le mode d'itération peut avoir des conséquences sur le comportement dynamique du réseau.

**Définition 1.22** (Robustesse)
*La* robustesse *d'un système peut être définie comme sa capacité à maintenir sa fonction, malgré des perturbations d'ordre structurel.*

La robustesse, telle que nous la définissons ici, peut être associée à la notion de stabilité structurelle introduite par Thom [73] dès le début des années 1970, qui est peut-être plus parlante.

D'un point de vue biologique, cette notion de robustesse est très importante. Supposons qu'un réseau de régulation soit connu pour être associé à une fonction biologique cyclique (comme le réseau de l'exemple 1.4). Si une modification de l'ordre dans lequel les gènes de ce réseau sont exprimés implique un changement du cycle limite atteint par le réseau, voire une disparition de ce cycle limite, alors, le réseau ne peut plus remplir la même fonction, voire, il peut ne plus être fonctionnel du tout. Inversement, le réseau de régulation peut être impliqué dans des processus pour lesquels une oscillation stable de l'expression des gènes n'est pas souhaitable, et dans ce cas, une apparition d'un cycle limite lorsque l'ordre d'expression des gènes est perturbé peut être problématique. En général, la nature est robuste et il existe des pare-feu qui permettent de limiter ou d'empêcher les perturbations aux conséquences importantes. Ainsi, on peut imaginer que, si les réseaux d'automates qui modélisent les réseaux de régulation ne sont pas robustes aux perturbations de leur mode d'itération, c'est qu'il existe un autre niveau dans la régulation génétique que les simples interactions de type activation/inhibition. Ce niveau garantirait la robustesse des réseaux de régulation génétique en imposant, en quelque sorte, une synchronisation de l'expression de

certains gènes.

*Exemple* 1.5 *:* Reprenons le réseau lié à la morphogénèse de la plume chez le poulet présenté dans l'exemple 1.4. Dans ce réseau, on considère que *Wnt2* est toujours actif (c'est un noeud source dont le seuil est négatif). Ainsi, on ne s'occupe pas de cet élément et on se concentre sur le réseau constitué des 3 autres éléments : *BMP7*, *BMP2* et la *Follistatine*. Le réseau d'automates booléens à seuil qui modélise ce réseau de régulation est alors défini par la matrice d'interaction $\mathbf{W}$ et le vecteur seuil $\boldsymbol{\theta}$ suivants (l'ordre des gènes est conservé) :

$$\mathbf{W} = \begin{pmatrix} 0 & -1 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & 0 \end{pmatrix} \quad \boldsymbol{\theta} = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}.$$

On remarque que le seuil associé au gène *BMP7* est passé de 0 à -1. Ceci nous permet de modéliser l'activation constante du gène *Wnt2* sur le gène *BMP7*. Comme *Wnt2* n'est plus présent dans le réseau réduit à 3 éléments, on échange le poids d'interaction de *Wnt2* sur *BMP7* par une décrémentation du seuil de *BMP7*.

Nous avons annoncé dans l'exemple 1.4 que ce RABS acceptait un unique cycle limite et pas de point fixe. C'est vrai quel que soit le mode d'itération blocs-séquentiel choisi. En revanche, le cycle limite change en fonction du mode d'itération, comme le montre le tableau 1.2. D'après la dernière ligne du tableau par exemple, le réseau, itéré avec le mode parallèle, atteint un cycle noté $[000, 100, 111, 010]$. Cette notation signifie que le cycle est constitué de 4 configurations du réseau. La configuration notée 100 indique que le gène 1 (ici c'est *BMP7*) est actif, alors que les 2 autres gènes sont inactifs.

Comme on associe les cycles limites du réseau d'automates booléens à la fonction remplie par le réseau de régulation génétique, le fait que les cycles limites soient modifiés par un changement de mode d'itération indique qu'un changement de synchronisme dans le mécanisme lié au processus d'expression des gènes a pour effet de changer la fonction du réseau. Pour que cette fonction soit conservée et que les spots puissent se mettre en place afin que le poulet ait des plumes qui poussent, il faut qu'il existe un dispositif permettant de synchroniser l'expression des 3 gènes d'une certaine manière.

Nous avons restreint notre étude aux conséquences de perturbations des modes d'itération sur les attracteurs des réseaux d'automates, et plus particulièrement, aux réseaux d'automates booléens à seuil. Nous ne traitons pas des phases transitoires, considérant que les attracteurs sont déjà de bons indicateurs du comportement dynamique global des réseaux. Nous pouvons définir 4 classes de comportement pour les réseaux d'automates, en fonction de leur robustesse au changement de modes d'itération blocs-séquentiels :

- la première classe de comportement englobe tous les réseaux qui n'ont que des cycles limites, quel que soit le mode d'itération avec lequel ils sont itérés ; ces réseaux n'ont jamais de points fixes ; on appelle cette classe, la classe "Cycle", et on la note *Cy* ;

TAB. 1.2 – Correspondance entre le mode d'itération blocs-séquentiel avec lequel le RABS de l'exemple 1.5 est itéré, et le cycle limite atteint par le réseau

| Mode d'itération | Cycle limite |
|---|---|
| $(1)(2)(3)$ | $[000, 110]$ |
| $(1)(3)(2)$ | $[000, 111]$ |
| $(1)(2,3)$ | $[000, 111]$ |
| $(2)(1)(3)$ | $[010, 101]$ |
| $(2)(3)(1)$ | $[010, 100]$ |
| $(2)(1,3)$ | $[010, 100]$ |
| $(3)(1)(2)$ | $[000, 110]$ |
| $(3)(2)(1)$ | $[011, 100]$ |
| $(3)(1,2)$ | $[000, 100, 011]$ |
| $(1,2)(3)$ | $[000, 101, 010]$ |
| $(1,3)(2)$ | $[000, 110]$ |
| $(2,3)(1)$ | $[011, 100]$ |
| $(1,2,3)$ | $[000, 100, 111, 010]$ |

- la deuxième classe de comportement est celle qui réunit tous les réseaux qui n'ont jamais de cycles limites, quel que soit le mode d'itération avec lequel ils sont itérés ; ces réseaux n'ont que des points fixes : on appelle donc cette classe, la classe "Fixe", et on la note *Fi* ;
- la troisième classe, que l'on appelle la classe "Mixte", et que l'on note *Mi*, est la classe de tous les réseaux qui ont à la fois au moins un point fixe et au moins un cycle limite, quel que soit le mode d'itération ;
- enfin, la quatrième classe de comportement regroupe tous les réseaux qui ont un comportement dynamique fortement dépendant du mode d'itération avec lequel ils sont itérés ; pour certains modes d'itération, ces réseaux n'ont que des points fixes, alors que pour les autres modes d'itération, ils ont également des cycles limites, en plus des points fixes ; on appelle cette classe, la classe "Evolution", et on la note *Ev*.

Nous avons vu précédemment que les points fixes d'un réseau étaient toujours les mêmes, quel que soit le mode d'itération avec lequel le réseau est itéré. Ainsi, les réseaux de la classe *Fi* ont toujours le même comportement dynamique. Si l'on devait classer les réseaux en terme de robustesse, la classe de réseaux *Fi* serait donc la classe des réseaux les plus robustes. A l'opposé, on pourrait dire que la classe de réseaux *Ev* est la classe des réseaux les moins robustes, puisque leur comportement dynamique est celui qui change le plus avec les modes d'itération. Mais c'est cette sensibilité aux perturbations des modes d'itération qui en fait également la classe de réseaux la plus intéressante à étudier. C'est sur l'étude des 4 classes que nous venons de définir que porte principalement notre travail.

Avant de clore ce chapitre, nous allons présenter les principaux résultats obte-

nus concernant la dynamique des réseaux d'automates booléens et la robustesse aux modes d'itération. Il n'est pas question ici de faire la liste exhaustive de tous les résultats de ce domaine, mais de présenter ceux qui nous semblent être d'intérêt majeur.

Les premiers travaux traitent de RABS dont la matrice d'interaction est symétrique, voire quasi-symétrique. Dans [74], il est prouvé qu'un RABS à matrice symétrique, itéré de façon parallèle ne peut admettre que des points fixes et des cycles limites. Dans [75], Goles montre que pour chaque mode d'itération blocs-séquentiel tel que la matrice du réseau est à diagonale bloc-dominante, le RABS n'admet que des points fixes. En particulier, les RABS à matrice symétrique dont tous les coefficients diagonaux sont positifs ou nuls n'admettent que des points fixes lorsqu'ils sont itérés de façon séquentielle. Ces résultats sont repris dans [72]. Enfin, dans [76, 77], les auteurs introduisent la notion d'énergie pour les RABS itérés par mode d'itération blocs-séquentiel. Ils montrent notamment que les RABS à matrice quasi-symétrique itérés séquentiellement n'admettent que des points fixes. C'est aussi le cas des réseaux potentiels, dans lesquels l'énergie se comporte comme le potentiel d'un système dynamique continu gradient, le gradient étant ici discret [78].

Récemment, Salinas [79] a montré que, pour les réseaux d'automates booléens à $n$ éléments dont le graphe d'interaction ne comporte aucun circuit autre que les boucles et dont chaque noeud $i$ ne dépend que des noeuds $i$ à $n$, les dynamiques séquentielles et parallèle sont équivalentes (ce résultat avait déjà été montré par Tchuente [80]). Salinas a également montré que la réciproque était vraie pour des réseaux dont le graphe d'interaction ne contient aucune boucle (et toujours aucun circuit) : si les dynamiques séquentielles et parallèle sont identiques alors, tout noeud $i$ ne dépend que des noeuds $i$ à $n$. Enfin, Goles et Salinas [79, 81] ont démontré que pour tout réseau d'automates booléens dont le graphe d'interaction ne contient aucune boucle négative, aucun cycle limite obtenu en itérant le réseau de façon séquentielle ne peut être obtenu avec une itération parallèle, et inversement.

Dans ce chapitre, nous avons présenté les notions de base qui sous-tendent notre travail. Nous avons, en particulier, montré comment nous utilisons les réseaux d'automates booléens à seuil comme modèles de réseaux de régulation génétique. Nous avons également expliqué quelles sont les caractéristiques dynamiques de ces réseaux auxquelles nous nous intéressons. Nous allons, dans le chapitre suivant, entrer dans le vif du sujet, en présentant des résultats globaux sur le nombre de réseaux d'automates booléens à seuil que l'on peut obtenir en prenant en compte jusqu'à 7 éléments, ainsi que des algorithmes nous permettant d'étudier le comportement dynamique de ces réseaux.

# Chapitre 2

# Enumérations et algorithme pour la simulation

Dans ce chapitre, nous allons présenter l'énumération des réseaux d'automates booléens à seuil. L'intérêt de cette énumération est double.

Tout d'abord, en soi, cette information est intéressante, notamment dans le cadre de la modélisation de réseaux de régulation génétique. Connaître le nombre de RABS que l'on peut construire avec $n$ éléments, permet de connaître le nombre de possibilités de modèles dynamiques que l'on peut mettre en place pour un réseau de régulation génétique avec $n$ gènes. Dans ce cadre de modélisation, c'est surtout le nombre de fonctions à seuil qui est intéressant : en effet, sachant qu'un gène est par exemple activé par $i$ de ses voisins et inhibé par $j$ autres éléments, connaître le nombre de fonctions à seuil permettant de modéliser $i$ activations et $j$ inhibitions, nous permet de comprendre le nombre de configurations possibles qui existent entre ces $i$ activations et $j$ inhibitions, et donc le nombre de façons dont le gène peut être régulé.

Mais surtout, l'intérêt majeur de l'énumération des RABS est qu'elle nous permet de mettre en place un algorithme pour la simulation de la dynamique de ces réseaux. Cet algorithme, permettant de tirer au hasard un RABS et d'en simuler le comportement dynamique, sera utilisé dans le but de donner des statistiques générales sur le comportement dynamique des RABS. Ces résultats expérimentaux seront présentés dans le chapitre 3. Afin que ces statistiques ne soient pas biaisées, il faut nous assurer que, dans notre algorithme, le tirage au hasard d'un réseau suive bien une loi uniforme sur la population totale des RABS. La connaissance de cette population, au niveau de sa taille, comme de sa structure, est donc nécessaire.

Nous restreignons notre étude aux RABS à 7 éléments, et donc aux fonctions booléennes à seuil à 7 arguments. Tout d'abord, nous avons vu dans la partie 1.2.2 que les réseaux d'interaction génétique étaient constitués de modules de petite taille. Dans ce cadre, nous considérons que nous travaillons sur des modules de réseaux de grande taille. De plus, chaque réseau de taille supérieure à 7 est trop grand et donc trop lourd à simuler de façon exhaustive (nous avons vu

par exemple que le nombre de modes d'itération blocs-séquentiels croît de façon exponentielle avec le nombre de noeuds). On ne peut donc donner de statistique significative sur les classes de comportement dynamique que l'on a introduites dans la partie 1.3.2 pour ces réseaux qu'après des simulations très coûteuses en temps d'exécution.

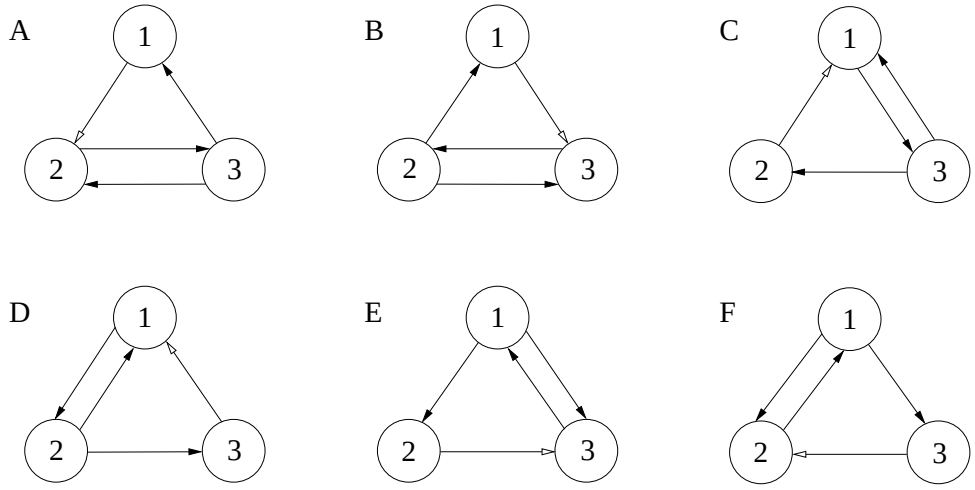



FIG. 2.1 – Graphes d'interaction correspondant à 6 réseaux de régulation composés de 3 gènes.

Les 6 graphes d'interaction de la figure 2.1 représentent 6 réseaux de régulation composés de 3 gènes (les réseaux sont désignés par les lettres A, B, C, D, E et F). On suppose que le poids d'interaction associé à chaque arc est égal à 1 si l'arc est positif, à -1 si l'arc est négatif. On suppose également que les seuils d'activation sont tous nuls. On peut remarquer, en changeant l'ordre des noeuds, que ces 6 réseaux sont en réalité 6 fois le même réseau. Par exemple, les gènes 1, 2 et 3 du réseau A sont respectivement régulés exactement de la même façon que les gènes 2, 1 et 3 du réseau C. Si l'on simule le comportement dynamique de ces 6 réseaux, on obtiendra donc 6 fois le même comportement, à l'ordre près des variables. Ainsi, les 6 RABS associés à ces 6 graphes d'interaction modélisent le même réseau de régulation génétique. Dans la suite, nous considérons donc que ces 6 réseaux sont un seul et même réseau que l'on ne comptera qu'une unique fois.

Nous venons de voir que, dans ce chapitre, nous allons travailler avec des permutations entre noeuds. Nous allons donc commencer par rappeler les notions relatives aux permutations d'un ensemble. Puis, nous présenterons l'énumération des fonctions à seuil ayant jusqu'à 7 arguments. A partir de ces fonctions à seuil, nous montrerons ensuite comment énumérer le nombre total de réseaux d'automates booléens à seuil composés de $n$ éléments ($n$ allant jusqu'à 7). Enfin, nous décrirons l'algorithme qui nous permet de tirer au hasard un RABS et d'en

simuler le comportement dynamique.

## 2.1 A propos des permutations

Dans cette partie, nous allons rappeler les notions relatives aux permutations d'un ensemble, qui nous seront utiles tout au long de ce chapitre.

On s'intéresse ici uniquement aux permutations de l'ensemble des $n$ noeuds d'un graphe, des $n$ éléments d'un réseau, ou encore des $n$ arguments d'une fonction booléenne. On donnera donc uniquement les définitions relatives aux permutations de l'ensemble $\{1, \cdots, n\}$.

**Définition 2.1** (Permutation)
*Une* permutation $\sigma$ *de l'ensemble* $\{1, \cdots, n\}$ *est une bijection de* $\{1, \cdots, n\}$ *dans lui-même.*

**Notation 2.1 :** *On note* $S_n$ *le groupe symétrique d'ordre* $n$*, c'est-à-dire le groupe des permutations de l'ensemble* $\{1, \cdots, n\}$*. On a* $\#S_n = n!$*.*
*On utilise la notation habituelle pour les permutations, c'est-à-dire :*
$\sigma \in S_n, \ \sigma = \begin{pmatrix} 1 & \cdots & n \\ \sigma(1) & \cdots & \sigma(n) \end{pmatrix}$.
*On note* $Id$ *la permutation identité :* $Id = \begin{pmatrix} 1 & \cdots & n \\ 1 & \cdots & n \end{pmatrix}$.

**Définition 2.2** (Support d'une permutation)
*On appelle* support d'une permutation $\sigma \in S_n$ *l'ensemble des éléments de* $\{1, \cdots, n\}$ *qui ne sont pas invariants par* $\sigma$*. Le support de* $\sigma$ *est donc l'ensemble défini par* $supp(\sigma) = \{x \in \{1, \cdots, n\} \ ; \ \sigma(x) \neq x\}$.

**Définition 2.3** (Cycle et permutation circulaire)
*Soit* $\sigma \in S_n$ *une permutation de* $\{1, \cdots, n\}$*, et* $k \in \mathbb{N}^*$ *un entier positif non nul. On dit que* $\sigma$ *est un* cycle *d'ordre* $k$ *si il existe une application injective* $i \mapsto x_i$ *de* $\{1, \cdots, k\}$ *dans* $\{1, \cdots, n\}$*, telle que* $\sigma$ *est définie par :*

$$\sigma(x) = \begin{cases} \sigma(x_j) = x_{j+1} & si \ j \leq k, \\ \sigma(x_k) = x_1, & \\ \sigma(x) = x & si \ x \notin \{x_1, \cdots, x_k\}. \end{cases}$$

$\sigma$ *est alors notée* $\sigma = (x_1, \cdots, x_k)$.
*Une* permutation circulaire *de* $\{1, \cdots, n\}$ *est un cycle de* $\{1, \cdots, n\}$ *dont le support est* $\{1, \cdots, n\}$ *tout entier.*

Les cycles nous permettent de donner une propriété très utile des permutations.

**Proposition 2.1** (Décomposition en produit de cycles)
*Toute permutation* $\sigma$ *de* $\{1, \cdots, n\}$ *se décompose de manière unique en produit de cycles de supports disjoints dont la réunion est* $\{1, \cdots, n\}$.

*Exemple* 2.1 *:* Soit $\sigma$ la permutation de $\{1,\cdots,5\}$ définie par $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 1 & 5 & 2 & 3 \end{pmatrix}$. On a $\sigma(1)=4, \sigma(4)=2$ et $\sigma(2)=1$, donc la restriction de $\sigma$ à $\{1,2,4\}$ est le cycle $(1,4,2)$. De même, la restriction de $\sigma$ à $\{3,5\}$ est le cycle $(3,5)$. La décomposition de $\sigma$ en produit de cycles de supports disjoints est donc : $\sigma = (1,4,2)(3,5)$.

Cette décomposition en produit de cycles à supports disjoints nous permet de noter plus simplement les permutations. Nous ne donnons pas ici de démonstration de cette proposition très classique.

Les permutations nous servent à changer l'ordre des éléments dans des vecteurs ou des matrices. Pour cela, nous utilisons les matrices de permutation.

**Définition 2.4** (Matrice de permutation)
*Soit $\sigma \in S_n$ une permutation de $\{1,\cdots,n\}$. La* matrice de permutation *associée à $\sigma$ est la matrice carrée de taille $n$, $\mathbf{P}_\sigma$, dont le terme général est*
$[\mathbf{P}_\sigma]_{ij} = \delta_{i,\sigma(j)} = \begin{cases} 1 & \textit{si } i=\sigma(j) \ \textit{(équivalent à } j=\sigma^{-1}(i)) \\ 0 & \textit{sinon} \end{cases}$.
*La transposée d'une matrice de permutation est également son inverse et on a* ${}^t\mathbf{P}_\sigma = \mathbf{P}_\sigma^{-1} = \mathbf{P}_{\sigma^{-1}}$.

Dans la matrice de permutation associée à $\sigma \in S_n$, le seul élément non nul et égal à 1 de la colonne $j$ se situe à la ligne $\sigma(j)$, $j \in \{1,\cdots,n\}$.

Afin d'envoyer l'élément $i$ du vecteur $\mathbf{X} \in \mathbb{R}^n$ à la place $\sigma(i)$, on multiplie le vecteur $\mathbf{X}$ par la matrice de permutation associée à $\sigma$. On note abusivement pour tout vecteur $\mathbf{X} \in \mathbb{R}^n$, $\sigma.\mathbf{X} = \mathbf{P}_\sigma.\mathbf{X}$. En notant $\mathbf{Y} = \sigma.\mathbf{X}$, on a $\forall i \in \{1,\cdots,n\}, y_i = x_{\sigma^{-1}(i)}$. En effet, $y_i = \sum_{j=1}^n \delta_{i,\sigma(j)}.x_j = \sum_{j=1}^n \delta_{\sigma^{-1}(i),j}.x_j = x_{\sigma^{-1}(i)}$. Ainsi, dans $\mathbf{Y}$, la coordonnée $\sigma(i)$ correspond bien à la coordonnée $i$ de $\mathbf{X}$.

De même pour les matrices : si $\mathbf{M} \in M_n(\mathbb{R})$ est une matrice carrée de taille $n$ à éléments dans $\mathbb{R}$ et que l'on veut envoyer l'élément de la ligne $i$ et de la colonne $j$ à la ligne $\sigma(i)$, colonne $\sigma(j)$, on multiplie à gauche par la matrice de permutation associée à $\sigma$ et à droite par la matrice de permutation associée à $\sigma^{-1}$. On note abusivement pour toute matrice $\mathbf{M} \in M_n(\mathbb{R})$, $\sigma.\mathbf{M} = \mathbf{P}_\sigma.\mathbf{M}.\mathbf{P}_{\sigma^{-1}}$. En notant $\mathbf{N} = \sigma.\mathbf{M}$, on a $\forall i \in \{1,\cdots,n\}, \forall j \in \{1,\cdots,n\}, n_{ij} = [\mathbf{N}]_{i,j} = [\mathbf{M}]_{\sigma^{-1}(i),\sigma^{-1}(j)} = m_{\sigma^{-1}(i),\sigma^{-1}(j)}$. En effet,
$n_{ij} = \sum_{k=1}^n \delta_{i,\sigma(k)}.\left(\sum_{l=1}^n m_{k,l}.\delta_{l,\sigma^{-1}(j)}\right) = \sum_{k=1}^n \delta_{\sigma^{-1}(i),k}.m_{k,\sigma^{-1}(j)} = m_{\sigma^{-1}(i),\sigma^{-1}(j)}$.
Ainsi, dans $\mathbf{N}$, l'élément situé ligne $\sigma(i)$, colonne $\sigma(j)$ correspond bien à l'élément situé ligne $i$, colonne $j$ de $\mathbf{M}$.

La matrice de permutation de la composition de la permutation $\sigma \in S_n$ par la permutation $\mu \in S_n$, notée $\mu.\sigma$, et qui remplace $i$ par $\mu(\sigma(i))$, $i \in \{1,\cdots,n\}$, est la matrice $\mathbf{P}_{\mu.\sigma} = \mathbf{P}_\mu \mathbf{P}_\sigma$. On vérifie facilement que la matrice $\mathbf{P}_{\mu.\sigma}$ est bien de terme général $[\mathbf{P}_{\mu.\sigma}]_{ij} = \delta_{i,\mu(\sigma(j))}$.

*Exemple* 2.2 *:* On reprend les réseaux A et D dont les graphes d'interaction sont donnés figure 2.1. Nous avons déjà supposé que les poids d'interaction appartenaient à l'ensemble $\{-1,0,1\}$ et que les seuils d'activation étaient nuls. Les matrices d'interaction correspondant à ces réseaux, $\mathbf{W}_A$ et $\mathbf{W}_D$ sont données

ci-dessous :

$$\mathbf{W}_A = \begin{pmatrix} 0 & 0 & 1 \\ -1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \mathbf{W}_D = \begin{pmatrix} 0 & 1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

On définit la permutation $\sigma = (1, 3, 2)$ (notation sous forme de cycle). $\sigma$ envoie 1 en 3, 3 en 2 et 2 en 1. La matrice de permutation associée à $\sigma$ est alors :

$$\mathbf{P}_\sigma = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

On a donc :

$$\sigma.\mathbf{W}_A = \mathbf{P}_\sigma.\mathbf{W}_A.\mathbf{P}_{\sigma^{-1}} = \mathbf{P}_\sigma.\mathbf{W}_A.{}^t\mathbf{P}_\sigma = \begin{pmatrix} 0 & 1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = \mathbf{W}_D.$$

Ainsi, en permutant les noeuds du réseau A selon la permutation $\sigma$, on obtient le réseau D. Ces 2 réseaux sont bien identiques à une permutation des noeuds près. Leur comportement dynamique est donc le même (toujours à une permutation des noeuds près). Ils modélisent tous deux le même réseau de régulation (exactement le même comportement dynamique, mais un étiquetage des gènes différent).

## 2.2 Enumération des fonctions booléennes à seuil

Les fonctions booléennes (linéaires) à seuil ont été introduites, dans les années 1940, par McCulloch et Pitts [20], pour leurs travaux sur les réseaux de neurones. Elles ont ensuite été l'objet d'une attention particulière pendant une douzaine d'années entre 1960 et 1972 [82, 83, 84]. A cette époque, les ingénieurs en électronique espéraient que l'usage de portes à seuil, à la place des traditionnelles portes logiques AND, OR et NOT, leur permettrait de fabriquer des circuits numériques moins complexes, comportant un nombre de composants moins élevé. A partir de 1972, ils se rendirent compte que la construction de tels circuits nécessitait une expertise technique hors de portée, et abandonnèrent les recherches sur ce sujet. Les travaux entrepris ont cependant permis de démontrer certains résultats mathématiques sur ces fonctions booléennes à seuil, notamment concernant leur énumération.

Le but de cette partie n'est pas d'établir de nouveaux résultats, ou encore de faire une simple énumération des fonctions booléennes à $n$ arguments. De telles énumérations ont déjà été réalisées [85, 86]. Nous souhaitons surtout porter l'attention du lecteur sur certaines propriétés de ces fonctions. Ces propriétés sont importantes dans le cadre de la modélisation des réseaux de régulation génétique par des réseaux d'automates booléens à seuil.

Comme nous l'avons vu dans le chapitre 1, nous pouvons totalement décrire un RABS par la donnée d'une matrice d'interaction et un vecteur seuil. De même, nous pouvons décrire une fonction booléenne à seuil par un vecteur.

**Définition 2.5** (Vecteur d'interaction)
*On appelle* vecteur d'interaction *tout vecteur* $\mathbf{V} = (w_1, \cdots, w_n, \theta) \in \mathbb{Z}^{n+1}$.
*Les* $w_i$ *sont appelés les poids d'interaction et* $\theta$*, le seuil. Afin de bien différencier les poids d'interaction et le seuil, on notera désormais* $\mathbf{V} = (w_1, \cdots, w_n)(\theta)$.
*On dit que* $\mathbf{V}$ *est le* vecteur d'interaction *associé à la fonction booléenne à seuil* $f : \{0,1\}^n \to \{0,1\}$ *si, et seulement si*

$$\forall \mathbf{X} \in \{0,1\}^n, \quad f(\mathbf{X}) = H\left(\sum_{i=1}^{n} \omega_i x_i - \theta\right)$$

*où* $H$ *est la fonction de Heaviside présentée dans la partie 1.2.2.*

Les fonctions booléennes à seuil nous servent de base pour énumérer et construire les RABS. Comme nous définissons tout RABS par une matrice d'interaction et un vecteur seuil, nous définissons toute fonction booléenne à seuil par un vecteur d'interaction. On peut donc voir les vecteurs d'interaction comme les lignes des systèmes "matrice d'interaction + vecteur seuil" servant à construire les RABS. Cette définition des fonctions booléennes à seuil par un vecteur d'interaction est naturelle dans le cadre de la modélisation des réseaux de régulation génétique par des RABS. En effet, la suite logique des étapes de construction du modèle à partir des données biologiques est la suivante (voir par exemple [44]) : on met tout d'abord en place un graphe d'interaction orienté et signé à partir des informations du type "le gène $j$ active (resp. inhibe) le gène $i$", puis on attribue à chaque arc $(j, i)$ une valeur correspondant à l'importance relative de $j$ dans la régulation du gène $i$, tout en définissant un seuil d'activation pour le gène $i$. Ainsi, cette dernière étape définit chaque fonction de transition locale du modèle sous forme de fonction booléenne à seuil, à partir des différents poids des arcs et du seuil d'activation.

**Notation 2.2 :** *On note* $FS_n$ *l'ensemble des fonctions booléennes à seuil à* $n$ *arguments.*

Nous montrons tout d'abord que toute fonction booléenne à seuil $f \in FS_n$ peut être associée à une région de $\mathbb{R}^{n+1}$. Il faut noter que, pour chaque fonction $f$, la région associée contient au moins un point de $\mathbb{Z}^{n+1}$ (voir par exemple [85]), et donc que l'on peut restreindre l'espace des poids d'interaction et des seuils à $\mathbb{Z}$. Considérons $P$ l'ensemble des parties de $\{1, \cdots, n\}$. A partir de tout élément $p$ de $P$, on peut définir l'équation d'un hyperplan de $\mathbb{R}^{n+1}$, sous la forme $(\sum_{i \in p} x_i) - x_{n+1} = 0$. L'ensemble des $2^n$ hyperplans de $\mathbb{R}^{n+1}$ définis de la sorte forme ce que l'on appelle un arrangement central d'hyperplans (tous les hyperplans passent par l'origine). Toute fonction booléenne à seuil de $FS_n$ correspond à l'une des régions convexes de $\mathbb{R}^{n+1}$ délimitées par ces hyperplans [87] (ces régions sont non réduites à un point). Ainsi, 2 points $\mathbf{A}$ et $\mathbf{B}$ de $\mathbb{R}^{n+1}$ contenus dans une même région définissent la même fonction booléenne à seuil $f$, c'est-à-dire :

$$\forall \mathbf{X} \in \{0,1\}^n, \;\; H(\sum_{i=1}^{n} a_i x_i - a_{n+1}) = H(\sum_{i=1}^{n} b_i x_i - b_{n+1}) = f(\mathbf{X}).$$

Dans la même région, il existe également un point $\mathbf{C} \in \mathbb{Z}^{n+1}$ qui définit $f$. Ainsi, chaque fonction booléenne à seuil $f \in FS_n$ peut être définie par un vecteur d'interaction $\mathbf{V} \in \mathbb{Z}^{n+1}$.

*Exemple* 2.3 *:* On considère les fonctions booléennes à seuil à 1 variable. Les 4 fonctions booléennes à 1 variable, notées $f_1$, $f_2$, $f_3$ et $f_4$ sont toutes des fonctions booléennes à seuil. Elles sont présentées dans le tableau suivant sous forme de table de vérité.

| $x$ | $f_1(x)$ | $f_2(x)$ | $f_3(x)$ | $f_4(x)$ |
|---|---|---|---|---|
| 0 | 0 | 1 | 0 | 1 |
| 1 | 0 | 1 | 1 | 0 |

Dans $\mathbb{R}^2$, les 2 équations suivantes définissent 2 droites qui divisent le plan $(0, w, \theta)$ en 4 régions comme sur la figure 2.2 :

$$\begin{cases} -\theta &= 0 \\ w - \theta &= 0 \end{cases}$$

Chacune des fonctions $f_i$ correspond à l'une de ces 4 cellules. On note $\mathcal{C}_i$ la cellule associée à la fonction $f_i$ (voir la figure 2.2, sur laquelle chaque demi-droite frontière a la même couleur que la cellule à laquelle elle appartient). On a :

$$\begin{aligned} \mathcal{C}_1 &= \{\mathbf{V} = (w, \theta) \in \mathbb{R}^2 \ ; \ w \leq \theta \text{ et } \theta \geq 0\} \\ \mathcal{C}_2 &= \{\mathbf{V} = (w, \theta) \in \mathbb{R}^2 \ ; \ w > \theta \text{ et } \theta < 0\} \\ \mathcal{C}_3 &= \{\mathbf{V} = (w, \theta) \in \mathbb{R}^2 \ ; \ w > \theta \text{ et } \theta \geq 0\} \\ \mathcal{C}_4 &= \{\mathbf{V} = (w, \theta) \in \mathbb{R}^2 \ ; \ w \leq \theta \text{ et } \theta < 0\} \end{aligned}$$

Tout point de la région $\mathcal{C}_i$ à coordonnées entières relatives définit donc un vecteur d'interaction associé à la fonction $f_i$. Dans la région $\mathcal{C}_1$, on peut par exemple choisir les 3 vecteurs d'interaction $\mathbf{V}_1 = (0, 0)$, $\mathbf{V}'_1 = (1, 1)$, et $\tilde{\mathbf{V}}_1 = (-1, 0)$, tous trois associés à la fonction booléenne à seuil $f_1$.

Cette caractérisation des fonctions booléennes à seuil de $FS_n$ par des régions de l'espace $\mathbb{R}^{n+1}$ montre que l'on a, pour chaque fonction $f \in FS_n$, une infinité de vecteurs d'interaction associés. Comme nous voulons énumérer et construire toutes les fonctions booléennes à seuil ayant jusqu'à $n$ arguments, il nous faut choisir un candidat parmi tous les vecteurs d'interaction associés.

Premièrement, ce candidat doit nous permettre de construire des RABS dont les graphes d'interaction sont de structure quasi-minimale (nous avons introduit cette notion à la suite de la définition 1.6). Ainsi, si le RABS modélise un réseau de régulation génétique, ceci nous garantit que chaque arc $(j, i)$ du graphe d'interaction représente une réelle influence du gène $j$ dans la régulation du gène $i$.

*Exemple* 2.4 *:* Les vecteurs $\mathbf{V}'_1 = (1, 1)$ et $\tilde{\mathbf{V}}_1 = (-1, 0)$ sont tous les deux des vecteurs associés à la fonction $f_1$ de l'exemple 2.3. On suppose que $f_1$ est utilisée

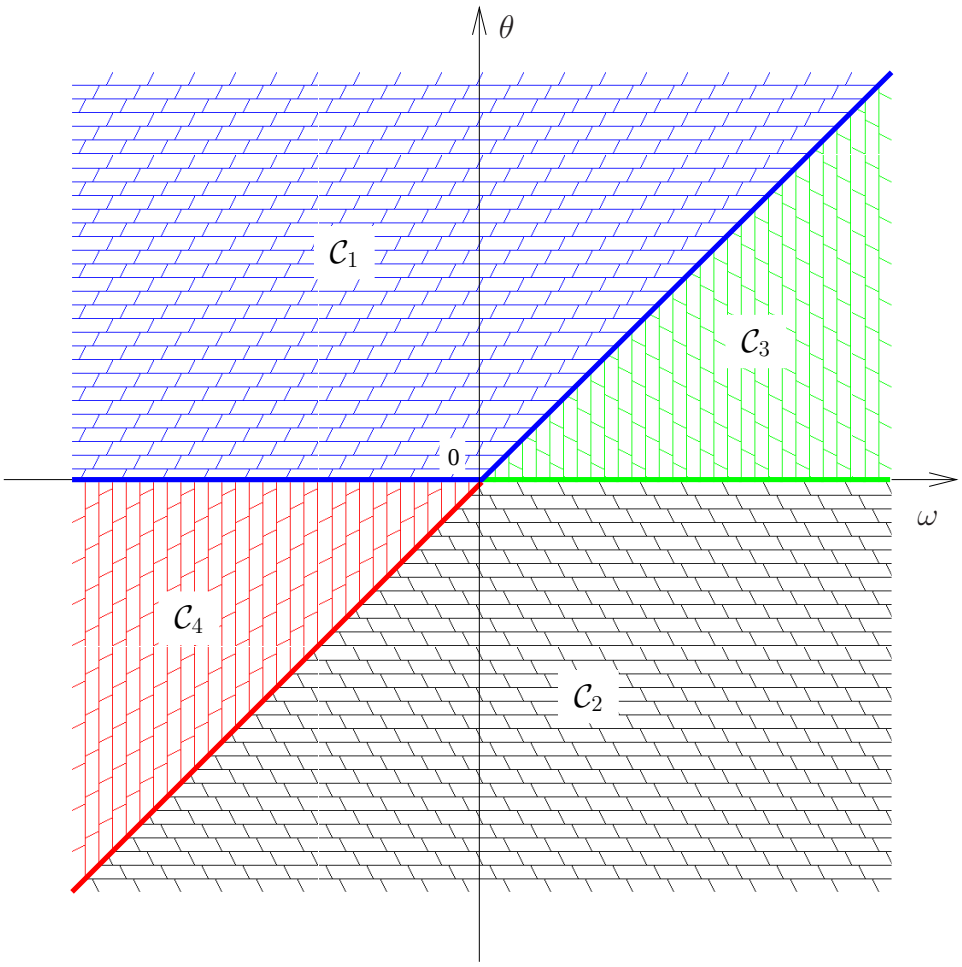



FIG. 2.2 – Division de $\mathbb{R}^2$ en 4 régions par les droites d'équation $\theta = 0$ et $\theta = w$.

dans un RABS comme fonction de transition locale associée à un élément $i$ ayant un voisin $j$ dans le réseau. Si $f_1$ est représentée par le vecteur d'interaction $\mathbf{V}'_1$ (resp. $\tilde{\mathbf{V}}_1$), alors, en observant le graphe d'interaction du réseau, on peut conclure que le gène $j$ est activateur (resp. inhibiteur) du gène $i$. Nous rappelons en effet que le graphe d'interaction d'un RABS est défini à partir de sa matrice d'interaction. Or, nous avons vu que $f_1$ est la fonction constante nulle. Donc $j$ n'a, en réalité, aucune influence sur $i$. Ainsi, ni $\mathbf{V}'_1$, ni $\tilde{\mathbf{V}}_1$ ne sont de bons candidats pour représenter la fonction booléenne à seuil $f_1$. On choisira donc de préférence $\mathbf{V}_1 = (0, 0)$ pour représenter $f_1$.

**Notation 2.3 :** *Soit* $\mathbf{X} \in \{0,1\}^n$ *un vecteur booléen, et* $i \in \{1, \cdots, n\}$ *un entier. On note* $\overline{\mathbf{X}}^i$ *le vecteur booléen* $\mathbf{Y} \in \{0,1\}^n$ *qui vérifie :*

$$y_j = \begin{cases} x_j & si \quad j \neq i \\ 1 - x_j & si \quad j = i \end{cases}$$

**Définition 2.6** (Vecteur d'interaction réduit)
*Soit* $\mathbf{V} \in \mathbb{Z}^{n+1}$*, un vecteur d'interaction associé à une fonction* $f \in FS_n$*. On dit que* $\mathbf{V} = (w_1, \cdots, w_n)(\theta)$ *est un* vecteur d'interaction réduit *si, et seulement si :*

$$\forall i \in \{1, \cdots, n\}, \;\; w_i \neq 0 \Longleftrightarrow \exists \mathbf{X} \in \{0,1\}^n \; / \; f(\mathbf{X}) = 1 - f(\overline{\mathbf{X}}^i).$$

Les fonctions booléennes à seuil étant des fonctions définies par signe, si $w_i$, poids d'interaction d'un vecteur d'interaction réduit, est non nul, son signe n'est pas ambigu. Ainsi, pour toute fonction $f \in FS_n$, il existe au moins un vecteur d'interaction associé qui soit réduit.

En définissant la fonction booléenne à seuil $f$ par un vecteur d'interaction réduit, on garantit que si $f$ modélise la régulation d'un gène $i$ dans un réseau de régulation génétique, alors tout arc $(j, i)$ du graphe d'interaction obtenu à partir de la matrice d'interaction, modélise une action réelle du gène $j$ sur le gène $i$. L'exemple décrit dans la partie 3.3 illustre un cas de modélisation de réseau de régulation génétique où les fonctions de transition locales, sous forme de fonctions booléennes à seuil, ne sont pas définies par des vecteurs d'interaction réduits.

**Notation 2.4 :** *Soit* $f \in FS_n$ *une fonction booléenne à seuil, et* $\mathbf{V} = (w_1, \cdots, w_n)(\theta) \in \mathbb{Z}^{n+1}$ *un vecteur d'interaction réduit associé à* $f$*.*
*On note* $I_f^+ = \{i \in \{1, \cdots, n\} \; ; \; w_i > 0\}$ *l'ensemble des indices des poids d'interaction strictement positifs de* $\mathbf{V}$*. On note* $k_f^+ = \#I_f^+$ *le nombre de ces indices.*
*On note* $I_f^- = \{i \in \{1, \cdots, n\} \; ; \; w_i < 0\}$ *l'ensemble des indices des poids d'interaction strictement négatifs de* $\mathbf{V}$*. On note* $k_f^- = \#I_f^-$ *le nombre de ces indices.*
*On dit que* $f$ *est une fonction à exactement* $k_f = k_f^+ + k_f^-$ *arguments, avec* $k_f^+$ *(resp.* $k_f^-$*) activations (resp. inhibitions).*

La définition des vecteurs d'interaction réduits nous garantit que les ensembles $I_f^+$ et $I_f^-$ ne dépendent pas du choix du vecteur $\mathbf{V}$.

**Proposition 2.2**
*Soit* $f \in FS_n$ *une fonction booléenne à seuil non constante, et* $\mathbf{V} = (w_1, \cdots, w_n)(\theta) \in \mathbb{Z}^{n+1}$ *un vecteur d'interaction réduit associé à* $f$*.*
*On note* $S^+ = \sum_{i \in I_f^+} w_i$ *et* $S^- = \sum_{i \in I_f^-} w_i$*. Si* $I_f^+$ *(resp.* $I_f^-$*) est vide, on prend* $S^+ = 0$ *(resp.* $S^- = 0$*).*
*On a alors :*

$$S^- \leq \theta < S^+.$$

*Preuve.* Remarquons que pour tout $\mathbf{X} \in \{0,1\}^n$, on a $S^- \leq \sum_{i=0}^n w_i x_i \leq S^+$. Ainsi, supposons que $\theta < S^-$. On a alors $H(S^- - \theta) = 1$ et donc $\forall X \in \{0,1\}^n$, $f(\mathbf{X}) = H\left(\sum_{i=0}^n w_i x_i - \theta\right) = 1$, ce qui est une contradiction avec le fait que $f$ n'est pas constante.
De même, en supposant que $\theta \geq S^+$, on trouve que $f$ est constante nulle. □

Les vecteurs d'interaction réduits ont donc leur seuil d'activation borné par des fonctions de leurs poids d'interaction.

Nous avons vu dans la partie 1.2.2 que, dans le cadre de la modélisation d'un réseau de régulation génétique par un RABS, la valeur absolue du poids d'interaction $w_{ij}$ représente l'importance du gène $j$ dans la régulation du gène $i$. Ainsi, si l'on veut définir une fonction booléenne à seuil à partir d'un vecteur d'interaction réduit, il est nécessaire que les différences entre les valeurs absolues des poids du vecteur, traduisent une différence d'importance entre les variables passées en arguments de la fonction.

*Exemple* 2.5 *:* On considère les 2 vecteurs d'interaction réduits de $\mathbb{Z}^4$ suivants : $\mathbf{V} = (1,1,2)(1)$ et $\mathbf{V}' = (1,2,3)(2)$. Ces 2 vecteurs d'interaction sont associés à la même fonction booléenne à seuil $f \in FS_3$ définie par la table de vérité présentée dans le tableau suivant ($f$ est une fonction à exactement 3 arguments avec 3 activations).

| $\mathbf{X}$ | 000 | 001 | 010 | 011 | 100 | 101 | 110 | 111 |
|---|---|---|---|---|---|---|---|---|
| $f(\mathbf{X})$ | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |

Si l'on définit $f$ à partir de $\mathbf{V}'$ et que l'on suppose que $f$ modélise la régulation d'un gène $i$ par 3 gènes notés 1, 2 et 3 (dans l'ordre des poids d'interaction de $\mathbf{V}'$), alors, l'effet du gène 2 dans la régulation de $i$ semble être plus important que l'effet du gène 1. D'après la table de vérité, ce n'est pas le cas. On peut permuter les gènes 1 et 2 sans que la régulation de $i$ n'en soit changée. $\mathbf{V}'$ est donc un mauvais candidat pour définir $f$, on lui préférera $\mathbf{V}$.

**Notation 2.5 :** *On considère $n \geq 2$. Soit $\mathbf{X} \in \{0,1\}^n$ un vecteur booléen, et $i,j \in \{1,\cdots,n\}, i \neq j$ deux entiers distincts. On note $\mathbf{X}^{+i}_{-j}$ le vecteur booléen $\mathbf{Y} \in \{0,1\}^n$ qui vérifie :*

$$y_k = \begin{cases} x_k & si \quad k \neq i, k \neq j \\ 1 & si \quad k = i \\ 0 & si \quad k = j \end{cases}$$

**Définition 2.7** (Vecteur d'interaction minimal)
*Soit $n \geq 2$. Soit $\mathbf{V} \in \mathbb{Z}^{n+1}$, un vecteur d'interaction réduit associé à une fonction $f \in FS_n$. On dit que $\mathbf{V} = (w_1,\cdots,w_n)(\theta)$ est un* vecteur d'interaction minimal *si, et seulement si :*

$$si\ k_f^+ > 1,\ \forall i,j \in I_f^+, i \neq j,\ \ w_i > w_j \Longleftrightarrow \exists \mathbf{X} \in \{0,1\}^n\ /\ f(\mathbf{X}^{+i}_{-j}) = 1 - f(\mathbf{X}^{+j}_{-i}) = 1$$
$$si\ k_f^- > 1,\ \forall i,j \in I_f^-, i \neq j,\ \ |w_i| > |w_j| \Longleftrightarrow \exists \mathbf{X} \in \{0,1\}^n\ /\ f(\mathbf{X}^{+i}_{-j}) = 1 - f(\mathbf{X}^{+j}_{-i}) = 0.$$

Par extension, si $f$ est une fonction booléenne à 1 variable, alors tout vecteur d'interaction réduit associé à $f$ est un vecteur d'interaction minimal. Le vecteur $\mathbf{V}'$ de l'exemple 2.5 n'est pas un vecteur minimal. Remarquons que si $\mathbf{V} = (w_1,\cdots,w_n)(\theta) \in \mathbb{Z}^{n+1}$ est un vecteur d'interaction minimal associé à la fonction booléenne à seuil $f \in FS_n$, alors, pour tout $k \in \mathbb{N}^*$, $k\mathbf{V} = (kw_1,\cdots,kw_n)(k\theta) \in \mathbb{Z}^{n+1}$ est également un vecteur d'interaction minimal associé à $f$ ; le vecteur d'interaction $\mathbf{V}^- = (-w_1,\cdots,-w_n)(-\theta-1)$ est, quant à lui, vecteur d'interaction minimal de la fonction $1-f$.

Les vecteurs d'interaction minimaux sont de bons candidats pour définir les fonctions booléennes à seuil. Si l'on utilise ces vecteurs d'interaction pour construire un RABS modèle de réseau de régulation génétique, alors on est assuré que :

- le graphe d'interaction du réseau est de structure quasi-minimale,
- les éléments de la matrice d'interaction représentent la réalité des importances relatives dans la régulation de chaque gène.

Nous énumérons et générons les fonctions booléennes à seuil en les définissant par des vecteurs d'interaction minimaux. Cependant, nous ne générons pas tous les éléments de $FS_n$. En effet, nous regroupons sous forme de classe d'équivalence les fonctions booléennes à seuil. Toutes les fonctions d'une même classe représentent le même comportement.

*Exemple* 2.6 *:* On considère les réseaux $A$, $B$ et $C$ de la figure 2.1. On s'intéresse aux fonctions de transition locale $f_2^A$, $f_3^B$ et $f_1^C$, où $f_i^R$ est la fonction de transition locale de l'élément $i$ du réseau $R$. Ces 3 fonctions booléennes à seuil sont décrites par la table de vérité donnée dans le tableau suivant.

| $\mathbf{X}$ | 000 | 001 | 010 | 011 | 100 | 101 | 110 | 111 |
|---|---|---|---|---|---|---|---|---|
| $f_2^A(\mathbf{X})$ | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| $f_3^B(\mathbf{X})$ | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| $f_1^C(\mathbf{X})$ | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |

Ces 3 fonctions sont bien différentes, pourtant elles représentent le même comportement. En effet, nous avons déjà vu que les 6 réseaux de la figure 2.1 sont en fait un seul et même réseau, dont les noeuds ont été renommés. Notamment, en appliquant la permutation $\sigma = (1,2)(3)$ au réseau $A$, on obtient le réseau $C$. De même, on a, pour tout $\mathbf{X} \in \{0,1\}^n$, $f_2^A(\sigma.\mathbf{X}) = f_1^C(\mathbf{X})$.

Nous définissons une relation d'équivalence sur les fonctions booléennes à seuil qui nous permet de réunir dans une même classe d'équivalence toutes les fonctions qui sont identiques, à une permutation près de leurs arguments.

**Définition 2.8** (Relation d'équivalence sur $FS_n$)
*Soient $f, g \in FS_n$ deux fonctions booléennes à seuil. On dit que $f$ et $g$ sont équivalentes selon $\mathcal{R}$ si, et seulement si :*

$$\exists \sigma \in S_n \ / \ \forall \mathbf{X} \in \{0,1\}^n, \quad f(\mathbf{X}) = g(\sigma.\mathbf{X})$$

*On note cette équivalence $f \ \mathcal{R} \ g$.*
*On peut alors définir la classe d'équivalence de tout élément $f$ de $FS_n$ (notée $\mathcal{R}(f)$) comme l'ensemble des éléments de $FS_n$ qui sont équivalents à $f$ selon $\mathcal{R}$. $\mathcal{R}(f) = \{g \in FS_n \ ; \ f \ \mathcal{R} \ g\}$. On réduit chaque classe d'équivalence afin de ne conserver que les éléments 2 à 2 distincts.*
*L'ensemble des classes d'équivalence de $FS_n$ selon $\mathcal{R}$ est l'ensemble $FS_n$ quotienté par $\mathcal{R}$, noté $FS_n/\mathcal{R}$ : $FS_n/\mathcal{R} = \{\mathcal{R}(f) \ ; \ f \in FS_n\}$.*

D'après les propriétés de groupe de $S_n$, on montre facilement que $\mathcal{R}$ est bien une relation d'équivalence sur $FS_n$.

A chaque fonction d'une classe d'équivalence de $FS_n/\mathcal{R}$, correspond un ensemble de vecteurs d'interaction minimaux. Nous utilisons ces vecteurs d'interaction minimaux pour définir un critère de sélection de l'élément de chaque classe d'équivalence que nous allons choisir comme représentatif de la classe.

**Notation 2.6 :** *Soit $f \in FS_n$ une fonction booléenne à seuil. On note $E_f \subset \mathbb{Z}^{n+1}$ l'ensemble des vecteurs d'interaction minimaux associés à $f$.*
*Soit $f \in FS_n$ une fonction booléenne à seuil, et $\mathbf{V} = (w_1, \cdots, w_n)(\theta) \in E_f$.*
*On note $\sigma_f \in S_n$ l'une des permutations telle que, dans $\sigma_f.\mathbf{V}$,*

- *les $k_f^+$ poids strictement positifs sont aux $k_f^+$ première places, les $k_f^-$ poids strictement négatifs sont aux $k_f^-$ places suivantes, et enfin les $n - k_f$ poids nuls sont aux places restantes ;*
- *les poids strictement positifs sont ordonnés par ordre croissant, les poids strictement négatifs par ordre décroissant.*

*D'après cette définition, $\sigma_f$ n'est unique que si tous les poids de $\mathbf{V}$ sont différents.*

La définition des vecteurs d'interaction minimaux nous garantit que la permutation $\sigma_f$ ne dépend pas du choix du vecteur $\mathbf{V}$.

**Définition 2.9** (Fonction représentative)
*Soit $\mathcal{C} \in FS_n/\mathcal{R}$ une classe d'équivalence de fonctions booléennes à seuil, selon $\mathcal{R}$. $f \in \mathcal{C}$ est dite* fonction représentative *de $\mathcal{C}$ si, et seulement si : $\sigma_f = Id$.*

Il est facile de vérifier que pour chaque classe d'équivalence, la fonction représentative existe et est unique.

De façon informelle, la fonction représentative d'une classe d'équivalence est la fonction booléenne à seuil dont les vecteurs d'interaction minimaux vérifient : les (éventuels) poids d'interaction strictement positifs sont les premiers, classés par ordre croissant, ensuite, on a les (éventuels) poids d'interaction strictement négatifs, classés par ordre décroissant, et enfin les (éventuels) poids d'interaction nuls.

**Définition 2.10** (Vecteur d'interaction représentatif)
*Soit $\mathcal{C} \in FS_n/\mathcal{R}$ une classe d'équivalence de fonctions booléennes à seuil, selon $\mathcal{R}$, et $f_\mathcal{C}$ la fonction représentative de $\mathcal{C}$.*
*On définit la suite finie d'ensembles de vecteurs minimaux associés à $f_\mathcal{C}$, $(U^i_{f_\mathcal{C}})_{1 \leq i \leq k_{f_\mathcal{C}}+1}$ comme suit :*

$$U^1_{f_\mathcal{C}} = \left\{ \mathbf{V} = (w_1, \cdots, w_n)(\theta) \in E_{f_\mathcal{C}} \ ; \ |w_1| = \min_{\mathbf{V}' \in E_{f_\mathcal{C}}} (|w'_1|) \right\},$$

$$U^i_{f_\mathcal{C}} = \left\{ \mathbf{V} = (w_1, \cdots, w_n)(\theta) \in U^{i-1}_{f_\mathcal{C}} \ ; \ |w_i| = \min_{\mathbf{V}' \in U^{i-1}_{f_\mathcal{C}}} (|w'_i|) \right\}, \ 2 \leq i \leq k_{f_\mathcal{C}},$$

$$U^{k_{f_\mathcal{C}}+1}_{f_\mathcal{C}} = \left\{ \mathbf{V} = (w_1, \cdots, w_n)(\theta) \in U^{k_{f_\mathcal{C}}}_{f_\mathcal{C}} \ ; \ |\theta| = \min_{\mathbf{V}' \in U^{k_{f_\mathcal{C}}}_{f_\mathcal{C}}} (|\theta'|) \right\}.$$

*Par construction, on a, pour tout* $i \in \{2, \cdots, k_{f_\mathcal{C}} + 1\}$, $U^i_{f_\mathcal{C}} \subseteq U^{i-1}_{f_\mathcal{C}}$.
$U^{k_{f_\mathcal{C}}+1}_{f_\mathcal{C}}$ *contient un unique élément,* $\mathbf{V}_{f_\mathcal{C}}$*, vecteur minimal associé à la fonction* $f_\mathcal{C}$*.* $\mathbf{V}_{f_\mathcal{C}}$ *est appelé* vecteur d'interaction représentatif *de* $\mathcal{C}$*.*

De façon informelle, le vecteur d'interaction $\mathbf{V}$, représentatif d'une classe d'équivalence de fonctions booléennes à seuil, est le vecteur minimal associé à la fonction $f$, représentative de la classe, obtenu par le processus suivant :

- on sélectionne le ou les vecteurs minimaux associés à $f$ dont la valeur absolue du premier poids est minimale ;
- s'il y a plusieurs vecteurs, on sélectionne parmi ces vecteurs celui ou ceux dont la valeur absolue du deuxième poids est minimale ;
- tant que la sélection de vecteurs contient plusieurs éléments, on répète le processus de sélection, jusqu'à minimiser la valeur absolue du seuil d'activation.

On peut définir toute fonction de la classe d'équivalence $\mathcal{C}$ à partir du vecteur d'interaction représentatif de $\mathcal{C}$. En effet, par définition de la relation d'équivalence, pour toute fonction $f$ (à $n$ arguments) de la classe d'équivalence dont $\mathbf{V} = (w_1, \cdots, w_n)(\theta)$ est le vecteur représentatif, il existe une permutation $\sigma \in S_n$ telle que $\mathbf{V}_\sigma = (w_{\sigma^{-1}(1)}, \cdots, w_{\sigma^{-1}(n)})(\theta)$ est un vecteur minimal associé à $f$.

**Notation 2.7 :** *Soit* $\mathbf{V} = (w_1, \cdots, w_n)(\theta) \in \mathbb{Z}^{n+1}$ *un vecteur d'interaction.*
*On note* $I(\mathbf{V})$ *l'ensemble des poids d'interaction de* $\mathbf{V}$*, 2 à 2 distincts. Pour chaque poids d'interaction* $i \in I(\mathbf{V})$*, on note* $n(\mathbf{V}, i)$ *le nombre d'occurences de* $i$ *dans les poids de* $\mathbf{V}$*,* $n(\mathbf{V}, i) = \#\{w_j = i,\ j \in \{1, \cdots, n\}\}$.

**Définition 2.11** (Vecteurs d'interaction de structure identique)
*Soient* $\mathbf{V}^1, \mathbf{V}^2 \in \mathbb{Z}^{n+1}$ *deux vecteurs d'interaction.*
*On dit que* $\mathbf{V}^1$ *et* $\mathbf{V}^2$ *sont de* structure identique *si, et seulement si il existe une bijection* $\phi$ *de* $I(\mathbf{V}^1)$ *dans* $I(\mathbf{V}^2)$ *telle que :* $\forall i \in I(\mathbf{V}^1),\ n(\mathbf{V}^1, i) = n(\mathbf{V}^2, \phi(i))$.

*Exemple* 2.7 *:* Les vecteurs d'interaction $\mathbf{V}^1 = (1, 3, 1, -2)(3)$ et $\mathbf{V}^2 = (1, 2, -2, -2)(2)$ sont de structure identique. Chacun d'eux a 2 poids d'interaction égaux parmi les 4. Ici, on peut prendre $\phi$ définie par $\phi(1) = -2, \phi(3) = 1, \phi(-2) = 2$.

**Définition 2.12** (Partie génératrice)
*Soit* $\mathcal{C} \in FS_n/\mathcal{R}$ *une classe d'équivalence de fonctions booléennes à seuil, et* $\mathbf{V}_\mathcal{C} = (w_1, \cdots, w_n)(\theta) \in \mathbb{Z}^{n+1}$ *le vecteur représentatif de* $\mathcal{C}$*.*
*L'ensemble de vecteurs d'interaction 2 à 2 distincts*
$PG_\mathcal{C} = \{\mathbf{V}_\sigma = (w_{\sigma^{-1}(1)}, \cdots, w_{\sigma^{-1}(n)})(\theta),\ \ \sigma \in S_n\}$ *est appelé* partie génératrice *de la classe d'équivalence* $\mathcal{C}$*.*
*Abusivement, on dit que* $\mathbf{V}_\mathcal{C}$ *est le vecteur représentatif de* $PG_\mathcal{C}$*.*
$PG_\mathcal{C}$ *contient* $\frac{n!}{\prod_{i \in I(\mathbf{V}_\mathcal{C})} n(\mathbf{V}_\mathcal{C}, i)!}$ *vecteurs d'interaction, chacun d'eux étant vecteur d'interaction minimal d'une unique fonction de* $\mathcal{C}$*. A chaque fonction de* $\mathcal{C}$ *est associé un unique vecteur d'interaction de* $PG_\mathcal{C}$*.*
*On note* $Gen(FS_n/\mathcal{R})$ *l'ensemble* $Gen(FS_n/\mathcal{R}) = \{PG_\mathcal{C},\ \mathcal{C} \in FS_n/\mathcal{R}\}$.

*Exemple* 2.8 *:* On reprend la fonction booléenne à seuil $f_2^A$ de l'exemple 2.6. En conservant la notation de l'exemple 2.6, la fonction représentative de la classe

d'équivalence de $\mathcal{R}(f_2^A)$ est la fonction $f_2^F$ (fonction de transition locale de l'élément 2 du réseau $F$ de la figure 2.1). Le vecteur d'interaction représentatif de $\mathcal{R}(f_2^A)$ est le vecteur d'interaction $(1,-1,0)(0)$. La partie génératrice de $\mathcal{R}(f_2^1)$ est l'ensemble de vecteurs d'interaction
$\{(1,-1,0)(0),(1,0,-1)(0),(-1,1,0)(0),(-1,0,1)(0),(0,1,-1)(0),(0,-1,1)(0)\}$.
Comme les poids d'interaction du vecteur d'interaction représentatif sont 2 à 2 distincts, la partie génératrice contient $3! = 6$ éléments.

En générant tous les vecteurs d'interaction représentatifs des classes d'équivalence des fonctions booléennes à $n$ arguments, on génère un représentant de chaque élément de l'ensemble $Gen(FS_n/\mathcal{R})$. On peut alors construire simplement l'ensemble de toutes les fonctions booléennes à seuil à $n$ éléments, en permutant les poids d'interaction du vecteur représentatif de chaque classe d'équivalence.

On énumère le nombre de classes d'équivalence des fonctions booléennes à seuil à $n$ arguments (pour $n \leq 7$) en générant tous les vecteurs d'interaction représentatifs de ces classes par l'algorithme exhaustif suivant :

1. on récupère les vecteurs d'interaction représentatifs des classes d'équivalence des fonctions booléennes à seuil à $n-1$ arguments $\rightarrow \mathcal{EV}_n$ ;
2. on fixe $p_{max}$ =poids maximum des vecteurs de $\mathcal{EV}_n$ ;
3. on génère tous les vecteurs d'interaction dont les poids sont ordonnés comme pour un vecteur représentatif de classe d'équivalence, et sont compris entre $-p_{max}$ et $p_{max}$ (le seuil est fonction des poids comme pour un vecteur d'interaction minimal) $\rightarrow \mathcal{EV}_n^{temp}$ ;
4. pour chaque vecteur $\mathbf{V}$ de $\mathcal{EV}_n^{temp}$, on construit la table de la fonction booléenne à seuil à laquelle le vecteur est associé ; si celle-ci ne correspond à aucun vecteur de $\mathcal{EV}_n$, $\mathbf{V} \rightarrow \mathcal{EV}_n$ ; sinon, si $\mathbf{V}$ minimise le critère de sélection énoncé précédemment, alors on remplace le vecteur associé à la même fonction que $\mathbf{V}$ par $\mathbf{V}$ dans $\mathcal{EV}_n$ ;
5. tant que le nombre total de fonctions que l'on peut obtenir à partir de l'ensemble de vecteurs représentatifs $\mathcal{EV}_n$ est inférieur au nombre total de fonctions booléennes à seuil à $n$ arguments donné par [85] ou [86] (voir le tableau 2.1), on incrémente $p_{max}$ et on revient à l'étape 3.

Nous aurions pu implémenter des algorithmes plus sophistiqués, et efficaces, basés sur des méthodes de programmation linéaire. Cependant, du fait de la puissance de calcul mise à disposition avec le calculateur Healthphy (CIMENT - Grenoble), le temps global de programmation et d'exécution n'aurait pas été significativement plus court. L'algorithme présenté ci-dessus est fortement parallélisable. Le calcul a donc pu être lancé sur 16 processeurs simultanément et le temps d'exécution a été de quelques dizaines d'heures pour construire les vecteurs d'interaction représentatifs de toutes les classes d'équivalence des fonctions booléennes à seuil à 7 arguments.

Dans le tableau 2.1, on voit que la proportion de fonctions à seuil parmi les fonctions booléennes est très faible, dès que le nombre d'arguments $n$ dépasse 4. Rappelons que le nombre de fonctions booléennes à $n$ arguments est $B_n = 2^{2^n}$.

TAB. 2.1 – $\#FS_n$, nombre de fonctions booléennes à seuil à $n$ arguments [85, 86], $B_n$, nombre de fonction booléennes à $n$ arguments et $p_n = \frac{\#FS_n}{B_n}$, proportion de fonctions booléennes à seuil parmi les fonction booléennes à $n$ arguments, pour $n \leq 7$.

| $n$ | $\#FS_n$ | $B_n$ | $p_n$ |
|---|---|---|---|
| 0 | 2 | 2 | 1 |
| 1 | 4 | 4 | 1 |
| 2 | 14 | 16 | 0,88 |
| 3 | 104 | 256 | 0,41 |
| 4 | 1 882 | 65 536 | 0,03 |
| 5 | 94 572 | 4 294 967 296 | $2,2 \times 10^{-5}$ |
| 6 | 15 028 134 | $1,84 \times 10^{19}$ | $8,1 \times 10^{-13}$ |
| 7 | 8 378 070 864 | $3,40 \times 10^{38}$ | $2,5 \times 10^{-29}$ |

Ainsi, modéliser la régulation de gènes par des fonctions booléennes, sans se restreindre à des fonctions définies par signe, comporte le risque de décrire un très grand nombre de comportements qui ne correspondent pas à une réalité biologique où les interactions sont de type activation/inhibition.

Le nombre de classes d'équivalence des fonctions booléennes à seuil à $n$ arguments (pour $n \leq 7$) est donné dans le tableau 2.2. Nous rappelons que ce nombre est le nombre d'éléments de l'ensemble $FS_n/\mathcal{R}$, c'est-à-dire $\#FS_n/\mathcal{R}$. Le tableau 2.3 donne la répartition du nombre de classes d'équivalence des fonctions booléennes à seuil à exactement $n$ arguments, en fonction du nombre d'activations de ces fonctions.

TAB. 2.2 – $\#FS_n/\mathcal{R}$, nombre de classes d'équivalence des fonctions booléennes à seuil à $n$ arguments et $\#FS_n^{exact}/\mathcal{R}$, nombre de classes d'équivalence des fonctions booléennes à seuil ayant exactement $n$ arguments, pour $n \leq 7$.

| $n$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| $\#FS_n/\mathcal{R}$ | 2 | 4 | 10 | 34 | 178 | 1 720 | 38 456 | 2 490 634 |
| $\#FS_n^{exact}/\mathcal{R}$ | 2 | 2 | 6 | 24 | 144 | 1 542 | 36 736 | 2 452 178 |

D'après le tableau 2.2, on remarque tout d'abord que, pour $n \geq 4$, la proportion de classes d'équivalence des fonctions à exactement $n$ arguments parmi les classes d'équivalence des fonctions à $n$ arguments, est très importante. Cette propriété a la conséquence suivante : les graphes d'interaction des RABS construits à partir de l'ensemble de ces fonctions booléennes à seuil sont de connexité élevée, comme nous le verrons dans la partie 2.3. D'après le tableau 2.3, on note également que le nombre de classes d'équivalence des fonctions à exactement $n$ arguments et $a \leq n$ activations est le même que le nombre de classes d'équivalence

TAB. 2.3 – Répartition du nombre de classes d'équivalence des fonctions booléennes à seuil à exactement $n$ arguments, en fonction du nombre d'activations $a$ de ces fonctions, pour $n \leq 7$.

| | $a=0$ | $a=1$ | $a=2$ | $a=3$ | $a=4$ | $a=5$ | $a=6$ | $a=7$ |
|---|---|---|---|---|---|---|---|---|
| $n=0$ | 2 | - | - | - | - | - | - | - |
| $n=1$ | 1 | 1 | - | - | - | - | - | - |
| $n=2$ | 2 | 2 | 2 | - | - | - | - | - |
| $n=3$ | 5 | 7 | 7 | 5 | - | - | - | - |
| $n=4$ | 17 | 34 | 42 | 34 | 17 | - | - | - |
| $n=5$ | 92 | 264 | 415 | 415 | 264 | 92 | - | - |
| $n=6$ | 994 | 4 046 | 8 208 | 10 240 | 8 208 | 4 046 | 994 | - |
| $n=7$ | 28 262 | 156 756 | 407 353 | 633 718 | 633 718 | 407 353 | 156 756 | 28 262 |

des fonctions à exactement $n$ arguments et $a$ inhibitions. Ceci est la conséquence directe de la remarque faite après la définition 2.7 concernant les vecteurs minimaux. Enfin, on voit que le nombre de comportements possibles de régulation d'un gène est très rapidement élevé, même si ce gène n'est régulé que par un faible nombre d'autres éléments. Par exemple, si on sait d'après la littérature ou d'après les résultats d'expérience qu'un gène $i$ est régulé par exactement 4 éléments sous la forme de 2 activations et 2 inhibitions, alors, d'après le tableau 2.3, il y a 42 configurations possibles de ces activations et inhibitions pour réguler $i$, sans compter toutes les permutations de chacune des configurations. Ceci montre donc que, si l'on veut modéliser un réseau de régulation génétique, en ayant simplement les connaissances qui permettent de mettre en place le graphe d'interaction, l'emploi de l'outil informatique est obligatoire afin de tester toutes les possibilités. Même avec la puissance atteinte par les ordinateurs ces dernières années, la simulation de réseau de régulation s'avère être un problème très coûteux en temps de calcul.

Le tableau 2.4 présente tous les vecteurs représentatifs de classes d'équivalence des fonctions booléennes à seuil ayant 3 arguments. Nous mettons à disposition ces vecteurs dans le but d'aider toute personne désirant modéliser un réseau de régulation sous la forme d'un RABS. Les 144 vecteurs représentatifs de classes d'équivalence des fonctions booléennes à seuil ayant exactement 4 arguments sont également fournis dans les tableaux 2 et 3 donnés en annexe. Les vecteurs correspondant aux fonctions à 5, 6 et 7 arguments sont disponibles sur demande[1].

Dans cette partie, nous avons exposé les propriétés qui nous semblent importantes pour représenter les fonctions booléennes à seuil de la meilleure façon possible, dans le cadre de la modélisation de réseaux de régulation génétique sous forme de RABS. Nous avons montré que l'on pouvait décrire chacune de ces

[1] Contact : adrien.elena@imag.fr

TAB. 2.4 – Vecteurs représentatifs des classes d'équivalence des fonctions booléennes à seuil à exactement $n$ arguments, pour $0 \leq n \leq 3$.

| | 0 activation | 1 activation | 2 activations | 3 activations |
|---|---|---|---|---|
| $n = 0$ | $(0,0,0)(0)$<br>$(0,0,0)(-1)$ | -<br>- | -<br>- | -<br>- |
| $n = 1$ | $(-1,0,0)(-1)$ | $(1,0,0)(0)$ | - | - |
| $n = 2$ | $(-1,-1,0)(-2)$<br>$(-1,-1,0)(-1)$ | $(1,-1,0)(-1)$<br>$(1,-1,0)(0)$ | $(1,1,0)(0)$<br>$(1,1,0)(1)$ | -<br>- |
| $n = 3$ | $(-1,-1,-2)(-3)$<br>$(-1,-1,-2)(-2)$<br>$(-1,-1,-1)(-3)$<br>$(-1,-1,-1)(-2)$<br>$(-1,-1,-1)(-1)$ | $(1,-1,-2)(-2)$<br>$(1,-1,-2)(-1)$<br>$(1,-1,-1)(-2)$<br>$(1,-1,-1)(-1)$<br>$(1,-1,-1)(0)$<br>$(2,-1,-1)(-1)$<br>$(2,-1,-1)(0)$ | $(1,1,-2)(-1)$<br>$(1,1,-2)(0)$<br>$(1,1,-1)(-1)$<br>$(1,1,-1)(0)$<br>$(1,1,-1)(1)$<br>$(1,2,-1)(-1)$<br>$(1,2,-1)(1)$ | $(1,1,1)(0)$<br>$(1,1,1)(1)$<br>$(1,1,1)(2)$<br>$(1,1,2)(1)$<br>$(1,1,2)(2)$ |

fonctions par un vecteur d'interaction que l'on considère optimal. Dans la partie suivante, nous montrons comment on peut construire et énumérer les RABS à partir de ces vecteurs d'interaction.

## 2.3 Enumération des réseaux d'automates booléens à seuil

A partir des vecteurs d'interaction, introduits dans la partie précédente pour définir les fonctions booléennes à seuil, nous pouvons construire les matrices d'interaction et vecteurs seuils qui représentent les réseaux d'automates booléens à seuil. Chaque ligne d'un système "matrice d'interaction + vecteur seuil" est obtenue à partir d'un vecteur d'interaction. Il est nécessaire de comprendre cette étape de construction des réseaux à partir des vecteurs d'interaction, d'une part pour pouvoir compter ces réseaux, mais aussi, et surtout, afin de pouvoir définir une méthode nous permettant de tirer au hasard un réseau dans la population totale de tous les RABS de taille donnée. Cette méthode de tirage au sort doit nous garantir que tous les individus de la population ont la même probabilité d'être sélectionnés. Pour cela, il nous faut connaître la structure de la population des RABS.

**Notation 2.8 :** *On note $RS_n$ l'ensemble des automates booléens à seuil constitués de $n$ éléments.*

Dans la suite, on associe tout élément $e$ de $RS_n$ à la matrice d'interaction $\mathbf{W}^e$ et au vecteur seuil $\boldsymbol{\theta}^e$ qui définissent les fonctions de transition locale de $e$. On note abusivement $e = (\mathbf{W}^e, \boldsymbol{\theta}^e)$. On dit que la ligne $i \in \{1, \cdots, n\}$ de

$e$ est construite à partir du vecteur d'interaction $\mathbf{V} = (w_1, \cdots, w_n)(\theta)$ si, pour tout $j \in \{1, \cdots, n\}$, $[\mathbf{W}^e]_{ij} = w_j$ et $[\boldsymbol{\theta}]_i = \theta$. On remarque que, si l'on note $f_i^e$, $i \in \{1, \cdots, n\}$, la fonction de transition locale de l'élément $i$ pour le réseau $e$, alors la ligne $i$ est construite à partir d'un vecteur d'interaction associé à $f_i^e$.

*Exemple* 2.9 *:* On considère $e = (\mathbf{W}^e, \boldsymbol{\theta}^e) \in RS_3$ le réseau défini comme suit :

$$\mathbf{W}^e = \begin{pmatrix} 1 & 1 & 0 \\ 0 & -1 & 1 \\ 2 & 1 & -1 \end{pmatrix} \quad \boldsymbol{\theta}^e = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}$$

La ligne 1 de $e$ est construite à partir du vecteur d'interaction $(1, 1, 0)(0)$, la ligne 2 à partir de $(0, -1, 1)(-1)$ et la ligne 3 à partir de $(2, 1, -1)(1)$.

**Définition 2.13** (Réseaux construits à partir de parties génératrices)
*Soit $E_{PG} = \{PG_1, \cdots, PG_n\} \subset Gen(FS_n/\mathcal{R})^n$ un ensemble de n parties génératrices d'éléments de $FS_n/\mathcal{R}$.*
*On définit l'ensemble des* réseaux construits à partir de *$E_{PG}$, noté $E_{RC} \subset RS_n$ comme l'ensemble de tous les RABS dont chaque ligne est construite à partir d'un des vecteurs d'interaction de chaque $PG_i$, $i \in \{1, \cdots, n\}$ en respectant la contrainte suivante : pour chaque RABS construit, chaque $PG_i$ contribue à une unique ligne $\sigma(i)$, $\sigma \in S_n$.*
*On réduit $E_{RC}$ de manière à ne conserver que les éléments 2 à 2 distincts.*

A partir de tous les ensembles possibles de $n$ parties génératrices d'éléments de $FS_n/\mathcal{R}$, on construit ainsi tous les RABS de $RS_n$.

**Proposition 2.3**
*On note $n_i$ la taille de l'élément $PG_i$ de $E_{PG} = \{PG_1, \cdots PG_n\} \subset Gen(FS_n/\mathcal{R})^n$, $i \in \{1, \cdots, n\}$, et $N_{tot}$ le nombre total de réseaux distincts que l'on construit à partir de $E_{PG}$. On a :*

$$N_{tot} \leq n! \prod_{i=1}^{n} n_i.$$

*Ce majorant est atteint dans le cas où les $PG_i$, $i \in \{1, \cdots, n\}$, sont 2 à 2 distincts.*

*Preuve.* Pour chaque élément $\sigma$ de $S_n$, le nombre de réseaux que l'on peut construire à partir de $E_{PG}$ qui sont tels que la ligne $\sigma(i)$ est construite à partir d'un vecteur de $PG_i$, est $\prod_{i=1}^{n} n_i$ ($n_1$ choix pour la ligne $\sigma(1)$ et $n_2$ choix pour la ligne $\sigma(2)$ $\cdots$ $n_n$ choix pour la ligne $\sigma(n)$). Il y a $n!$ permutations de l'ensemble $\{1, \cdots, n\}$, donc on peut construire $n! \prod_{i=1}^{n} n_i$ réseaux.
Si les $PG_i$ sont 2 à 2 distincts, alors les réseaux construits sont également 2 à 2 distincts et donc $N_{tot} = n! \prod_{i=1}^{n} n_i$.
En revanche, s'il existe au moins 2 parties génératrices égales (on a donc $n \geq 2$) contenant au moins 2 vecteurs d'interaction, $N_{tot} < n! \prod_{i=1}^{n} n_i$. Considérons par exemple $PG_1 = PG_2$. On note $\mathbf{V}^1$ le vecteur d'interaction représentatif de $PG_1$. Comme $PG_1$ contient au moins 2 vecteurs d'interaction, $\mathbf{V}^1$ contient 2 poids d'interaction aux valeurs différentes. On suppose, sans perte de généralité, que ces 2

poids sont les 2 premiers. On note $\pi_{12} \in S_n$ la permutation dont le support est réduit à $\{1,2\}$ ($\pi_{12}$ permute 1 et 2 et laisse invariants les autres entiers inférieurs à $n$). On a donc $\mathbf{V}^1 \neq \pi_{12}.\mathbf{V}^1$. Pour tout ensemble de $n-2$ vecteurs d'interaction $\{\mathbf{V}^3, \cdots, \mathbf{V}^n\}$ avec $\mathbf{V}^i \in PG_i$, $i \in \{3, \cdots, n\}$, le réseau obtenu en choisissant $\mathbf{V}^1$ pour la première ligne, $\pi_{12}.\mathbf{V}^1$ pour la deuxième ligne et $\mathbf{V}^i$ pour la ligne $i$, $i \in \{3, \cdots, n\}$ est le même réseau que celui obtenu en choisissant $\pi_{12}.\mathbf{V}^1$ pour la première ligne, $\mathbf{V}^1$ pour la deuxième ligne, et $\mathbf{V}^i$ pour la ligne $i$, $i \in \{3, \cdots, n\}$. Dans ce cas, on a donc bien $N_{tot} < n! \prod_{i=1}^{n} n_i$. □

*Exemple* 2.10 *:* On considère $n = 2$, et les 2 vecteurs d'interaction représentatifs $\mathbf{V}_1 = (-1,0)(-1)$ et $\mathbf{V}_2 = (1,0)(0)$. $\mathbf{V}_1$ et $\mathbf{V}_2$ sont les vecteurs représentatifs des parties génératrices $PG_1$ et $PG_2$ suivantes : $PG_1 = \{(-1,0)(-1), (0,-1)(-1)\}$ et $PG_2 = \{(1,0)(0), (0,1)(0)\}$. A partir de ces 2 parties génératrices, on peut construire un ensemble de 8 RABS $N^k$, $1 \leq k \leq 8$ dont les matrices d'interaction $\mathbf{W}^k$ et vecteurs seuils $\boldsymbol{\theta}^k$ sont les suivants :

$$\left(\mathbf{W}^1 = \begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix}, \boldsymbol{\theta}^1 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}\right) \quad \left(\mathbf{W}^2 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \boldsymbol{\theta}^2 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}\right)$$
$$\left(\mathbf{W}^3 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \boldsymbol{\theta}^3 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}\right) \quad \left(\mathbf{W}^4 = \begin{pmatrix} 0 & -1 \\ 0 & 1 \end{pmatrix}, \boldsymbol{\theta}^4 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}\right)$$
$$\left(\mathbf{W}^5 = \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix}, \boldsymbol{\theta}^5 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}\right) \quad \left(\mathbf{W}^6 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \boldsymbol{\theta}^6 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}\right)$$
$$\left(\mathbf{W}^7 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \boldsymbol{\theta}^7 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}\right) \quad \left(\mathbf{W}^8 = \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix}, \boldsymbol{\theta}^8 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}\right)$$

Le but n'est pas de construire et compter tous les éléments de $RS_n$ pour un $n$ donné. Comme nous l'avons fait pour les fonctions booléennes à seuil, nous définissons une relation d'équivalence entre les RABS. Cette relation d'équivalence nous permet de regrouper dans une même classe tous les réseaux qui sont identiques à une permutation près de leurs éléments, comme les 6 réseaux de la figure 2.1, par exemple.

**Définition 2.14** (Relation d'équivalence sur $RS_n$)
*Soient $N^1, N^2 \in RS_n$ deux réseaux d'automates booléens à seuil comportant le même nombre n d'éléments. On note $f_i^k$ la fonction de transition locale de l'élément i pour le réseau $N^k$, $k \in \{1,2\}$. On dit que $N^1$ et $N^2$ sont équivalents selon $\tilde{\mathcal{R}}$ si, et seulement si :*

$$\exists \sigma \in S_n \ / \ \forall i \in \{1, \cdots, n\}, \forall \mathbf{X} \in \{0,1\}^n, \ \ f_i^1(\mathbf{X}) = f_{\sigma(i)}^2(\sigma.\mathbf{X})$$

*On note cette équivalence $N^1 \ \tilde{\mathcal{R}} \ N^2$.*
*On peut alors définir la classe d'équivalence de tout élément $N$ de $RS_n$ (notée $\tilde{\mathcal{R}}(N)$) comme l'ensemble des éléments de $RS_n$ qui sont équivalents à $N$ selon $\tilde{\mathcal{R}}$. $\tilde{\mathcal{R}}(N) = \{N' \in RS_n \ ; \ N \ \tilde{\mathcal{R}} \ N'\}$. On réduit chaque classe d'équivalence afin de ne conserver que les éléments 2 à 2 distincts.*
*L'ensemble des classes d'équivalence de $RS_n$ selon $\mathcal{R}$ est l'ensemble $RS_n$ quotienté par $\tilde{\mathcal{R}}$, noté $RS_n/\tilde{\mathcal{R}}$ : $RS_n/\tilde{\mathcal{R}} = \{\tilde{\mathcal{R}}(N) \ ; \ N \in RS_n\}$.*

D'après les propriétés de groupe de $S_n$, on montre facilement que $\tilde{\mathcal{R}}$ est bien une relation d'équivalence sur $RS_n$. Deux réseaux équivalents selon $\tilde{\mathcal{R}}$ sont en fait le même réseau dont on a réorganisé les éléments. Il est donc inutile de les considérer comme 2 réseaux différents. C'est la raison pour laquelle on énumère l'ensemble des classes d'équivalence de $RS_n$ selon $\tilde{\mathcal{R}}$ et non l'ensemble de tous les éléments de $RS_n$. Comme nous l'avons vu, nous définissons tout RABS par sa matrice d'interaction et son vecteur seuil, construits à partir des vecteurs d'interaction définis dans la partie précédente. Il est donc souhaitable de définir la notion d'équivalence entre 2 RABS à partir de leurs matrices d'interaction et de leurs vecteurs seuils.

On remarque tout d'abord que si les réseaux sont construits à partir de l'ensemble $Gen(FS_n/\mathcal{R})$, alors 2 RABS équivalents selon $\tilde{\mathcal{R}}$ sont construits à partir du même ensemble de $n$ parties génératrices (les fonctions de transition locale des 2 réseaux appartiennent aux mêmes classes d'équivalence de $FS_n/\mathcal{R}$). Pour la suite, on suppose donc que tout élément de $RS_n$ est obtenu par construction à partir de $Gen(FS_n/\mathcal{R})$.

**Proposition 2.4**
*Soient $N^1, N^2 \in RS_n$ deux réseaux d'automates booléens à seuil comportant le même nombre $n$ d'éléments.*
*On note $\mathbf{W}^k$ la matrice d'interaction et $\boldsymbol{\theta}^k$ le vecteur seuil associé au réseau $N^k$, $k \in \{1,2\}$.*
*On a l'équivalence suivante :*

$$N^1 \,\tilde{\mathcal{R}}\, N^2 \iff \exists \sigma \in S_n \,/\, \mathbf{W}^2 = \sigma.\mathbf{W}^1 \;\; et \;\; \boldsymbol{\theta}^2 = \sigma.\boldsymbol{\theta}^1$$

.

*Preuve.* On note $f_i^k$ la fonction de transition locale de l'élément $i$ pour le réseau $N^k$, $k \in \{1,2\}$.
Supposons que $\exists \sigma \in S_n \,/\, \mathbf{W}^2 = \sigma.\mathbf{W}^1$ et $\boldsymbol{\theta}^2 = \sigma.\boldsymbol{\theta}^1$. On a donc, pour tout $i,j \in \{1,\cdots,n\}$, $w^2_{ij} = w^1_{\sigma^{-1}(i)\sigma^{-1}(j)}$ et $\theta^2_i = \theta^1_{\sigma^{-1}(i)}$.
Soit $i \in \{1,\cdots,n\}$, et $\mathbf{X} \in \{0,1\}^n$. On a :

$$\begin{aligned}
f^2_{\sigma(i)}(\sigma.\mathbf{X}) &= H\left(\sum_{j=1}^{n} w^2_{\sigma(i)j} x_{\sigma^{-1}(j)} - \theta^2_{\sigma(i)}\right) \\
&= H\left(\sum_{j=1}^{n} w^1_{\sigma^{-1}(\sigma(i))\sigma^{-1}(j)} x_{\sigma^{-1}(j)} - \theta^1_{\sigma^{-1}(\sigma(i))}\right) \\
&= H\left(\sum_{j=1}^{n} w^1_{i\sigma^{-1}(j)} x_{\sigma^{-1}(j)} - \theta^1_i\right) \\
&= H\left(\sum_{j=1}^{n} w^1_{ij} x_j - \theta^1_i\right) \\
&= f^1_i(\mathbf{X}).
\end{aligned}$$

On a bien $\left(\exists \sigma \in S_n \ / \ \mathbf{W}^2 = \sigma.\mathbf{W}^1 \text{ et } \boldsymbol{\theta}^2 = \sigma.\boldsymbol{\theta}^1\right) \implies N^1 \ \tilde{\mathcal{R}} \ N^2$.

Supposons maintenant que $N^1 \ \tilde{\mathcal{R}} \ N^2$.
On note $\mathbf{V}_i^k = (w_{i1}^k, \cdots, w_{in}^k)(\theta_i^k)$ le vecteur minimal associé à la fonction de transition locale $f_i^k$, choisi pour construire cette fonction, $i \in \{1, \cdots, n\}, k \in \{1, 2\}$. La matrice d'interaction $\mathbf{W}^k$ est donc de terme général $w_{ij}^k$ et le vecteur seuil $\boldsymbol{\theta}^k$ de terme général $\theta_i^k$, $k \in \{1, 2\}$.
$N^1$ et $N^2$ étant équivalents selon $\tilde{\mathcal{R}}$, il existe une permutation $\sigma \in S_n$ telle que $\forall i \in \{1, \cdots, n\}, \forall \mathbf{X} \in \{0,1\}^n, \quad f_i^1(\mathbf{X}) = f_{\sigma(i)}^2(\sigma.\mathbf{X})$. Ainsi, pour tout $i \in \{1, \cdots, n\}$, les fonctions $f_i^1$ et $f_{\sigma(i)}^2$ sont équivalentes selon $\mathcal{R}$. Comme les matrices d'interaction et vecteurs seuils sont construits à partir des vecteurs représentatifs des éléments de $FS_n/\mathcal{R}$, il existe une permutation $\mu_i \in S_n$ telle que
$\mathbf{V}_{\sigma(i)}^2 = (w_{i\mu_i^{-1}(1)}^1, \cdots, w_{i\mu_i^{-1}(n)}^1)(\theta_i^1)$. Pour tout $\mathbf{X} \in \{0,1\}^n$, on a donc :

$$\begin{aligned} f_i^1(\mathbf{X}) &= H\left(\sum_{j=1}^{n} w_{ij}^1 x_j - \theta_i^1\right) \\ &= f_{\sigma(i)}^2(\sigma.\mathbf{X}) \\ &= H\left(\sum_{j=1}^{n} w_{i\mu_i^{-1}(j)}^1 x_{\sigma^{-1}(j)} - \theta_i^1\right) \end{aligned}$$

Par identification terme à terme, on trouve $\mu_i = \sigma$ pour tout $i \in \{1, \cdots, n\}$. On a donc $V_i^2 = (w_{\sigma^{-1}(i)\sigma^{-1}(1)}^1, \cdots, w_{\sigma^{-1}(i)\sigma^{-1}(n)}^1)(\theta_{\sigma^{-1}(i)}^1)$ pour tout $i \in \{1, \cdots, n\}$.
D'où $\mathbf{W}^2 = \sigma.\mathbf{W}^1$ et $\boldsymbol{\theta}^2 = \sigma.\boldsymbol{\theta}^1$. □

La relation d'équivalence entre RABS est donc également définie à partir des matrices d'interaction et vecteurs seuils.

A partir de l'ensemble $E_{PG}$ de $n$ parties génératrices d'éléments de $FS_n/\mathcal{R}$, on construit un ensemble $E_{RC} \subset RS_n$ de réseaux. Combien de classes d'équivalence de $RS_n/\tilde{\mathcal{R}}$ sont alors définies par l'ensemble $E_{RC}$ ? Si l'on peut répondre à cette question, alors, comme nous avons généré l'ensemble $Gen(FS_n/\mathcal{R})$ dans la partie précédente, on sait énumérer toutes les classes d'équivalence de RABS à $n$ éléments, et on connaît la structure de $RS_n/\tilde{\mathcal{R}}$.

*Exemple* 2.11 *:* On considère les 8 RABS construits dans l'exemple 2.10. Ces 8 réseaux définissent 4 classes d'équivalence de $RS_2/\tilde{\mathcal{R}}$.
En effet, en notant $\sigma = (1,2)$ (notation sous forme de cycle), on a $\mathbf{W}^8 = \sigma.\mathbf{W}^1$ et $\boldsymbol{\theta}^8 = \sigma\boldsymbol{\theta}^1$, donc $N^1 \ \tilde{\mathcal{R}} \ N^8$. On montre de même que, $N^2 \ \tilde{\mathcal{R}} \ N^6$, $N^3 \ \tilde{\mathcal{R}} \ N^7$ et $N^4 \ \tilde{\mathcal{R}} \ N^5$. Les 8 RABS définissent donc 4 classes d'équivalence.

On peut remarquer que, pour énumérer les classes d'équivalence de RABS, et générer un élément représentatif de chaque classe, à partir d'un ensemble de parties génératrices $E_{PG} = \{PG_1, \cdots, PG_n\} \subset Gen(FS_n/\mathcal{R})^n$, on peut limiter le nombre de réseaux construits à partir de $E_{PG}$ en ajoutant la contrainte suivante :

chaque partie génératrice $PG_i$, $i \in \{1, \cdots, n\}$, contribue à la ligne $i$ de chaque réseau construit. Avec cette contrainte, on ne peut pas construire tous les RABS de $RS_n$, mais on est assuré de générer au moins un élément de chaque classe d'équivalence de $RS_n/\tilde{\mathcal{R}}$. En effet, considérons un RABS $N \in RS_n$ construit à partir de $E_{PG}$ et tel que chaque ligne $i \in \{1, \cdots, n\}$ est obtenue à partir de la partie génératrice $PG_{\sigma(i)}$, $\sigma \in S_n, \sigma \neq Id$. Le réseau $N^\sigma$, dont la matrice d'interaction est $\mathbf{W}^\sigma = \sigma.\mathbf{W}$ et le vecteur seuil est $\boldsymbol{\theta}^\sigma = \sigma.\boldsymbol{\theta}$, est équivalent à $N$ selon $\tilde{\mathcal{R}}$, d'après la proposition 2.4. Par définition de $N^\sigma$, la ligne $i \in \{1, \cdots, n\}$ de $N^\sigma$ est construite à partir d'un vecteur d'interaction de $PG_i$, .

De cette façon, on construit donc jusqu'à $\prod_{i=1}^{n} n_i$ réseaux distincts ($n_i$ représente le nombre de vecteurs d'interaction de $PG_i$), au lieu de $n! \prod_{i=1}^{n} n_i$. On obtient donc un majorant du nombre de classes d'équivalence de $RS_n/\tilde{\mathcal{R}}$ définies par les réseaux construits à partir de $E_{PG}$, puisque ce nombre ne peut pas être supérieur au nombre de réseaux construits. Ce majorant est $\prod_{i=1}^{n} n_i$.

**Proposition 2.5**
*Soient $E_{PG} = \{PG_1, \cdots, PG_n\}$ et $\tilde{E}_{PG} = \{\tilde{PG}_1, \cdots, \tilde{PG}_n\}$ 2 ensembles de n parties génératrices d'éléments de $FS_n/\mathcal{R}$. On suppose que les éléments de $E_{PG}$ (resp. $\tilde{E}_{PG}$) ne sont pas 2 à 2 distincts.*
*On note $\mathbf{V}_i$ et $\tilde{\mathbf{V}}_i$ les vecteurs représentatifs de $PG_i$ et $\tilde{PG}_i$, $i \in \{1, \cdots, n\}$.*
*On note $N_{equiv}$ et $\tilde{N}_{equiv}$ les nombres de classes d'équivalence de $RS_n/\tilde{\mathcal{R}}$ définies par les réseaux construits à partir de $E_{PG}$ et $\tilde{E}_PG$.*

*On a alors :*

$$\begin{cases} \forall i \in \{1, \cdots, n\},\ \mathbf{V}_i \text{ et } \tilde{\mathbf{V}}_i \text{ sont de structure identique} \\ \forall i, j \in \{1, \cdots, n\},\ PG_i = PG_j \Leftrightarrow \tilde{PG}_i = \tilde{PG}_j \end{cases} \Rightarrow N_{equiv} = \tilde{N}_{equiv}$$

*On dit que $E_{PG}$ et $\tilde{E}_{PG}$ sont de structure générale identique.*

*Preuve.* Ce problème est un problème d'étiquetage des éléments des matrices d'interaction et vecteurs seuils. Les matrices obtenues à partir de $E_{PG}$ sont les mêmes que celle obtenues à partir de $\tilde{E}_{PG}$, à un changement d'étiquette de leurs éléments près. De même pour les vecteurs seuils.
On suppose, sans perte de généralité, que $PG_1 = PG_2$ et donc $\tilde{PG}_1 = \tilde{PG}_2$. On traite uniquement le cas $PG_1 \neq PG_i$, $2 < i \leq n$, $\tilde{PG}_1 \neq \tilde{PG}_i$, $2 < i \leq n$ et $PG_i = \tilde{PG}_i$, $2 < i \leq n$. La démonstration du cas général suit le même raisonnement, mais nécessite des notations plus lourdes.
On note $\phi$ la bijection qui permet de passer des poids de $\mathbf{V}_1$ à $\tilde{\mathbf{V}}_1$, comme dans la définition 2.11 des vecteurs d'interaction de structure identique. On considère que les vecteurs d'interaction sont rangés de la même manière dans $PG_1$ et dans $\tilde{PG}_1$ : en notant $\mathbf{V}^i = (w_1^i, \cdots, w_n^i)(\theta)$ (resp. $\tilde{\mathbf{V}}^i = (\tilde{w}_1^i, \cdots, \tilde{w}_1^i)(\tilde{\theta})$) l'élément $i$ de $PG_1$ (resp. $\tilde{PG}_1$), on a pour tout $j \in \{1, \cdots, n\}$, $\tilde{w}_j^i = \phi(w_j^i)$.
On considère 2 réseaux distincts $N^1 = (\mathbf{W}^1, \boldsymbol{\theta}^1)$ et $N^2 = (\mathbf{W}^2, \boldsymbol{\theta}^2)$ construits à partir de $E_{PG}$ : chaque ligne $i$ de $N^1$ et $N^2$ est un vecteur d'interaction de $PG_i$, $i \in \{1, \cdots, n\}$. On note $\tilde{N}^1 = (\tilde{\mathbf{W}}^1, \tilde{\boldsymbol{\theta}^1})$ (resp. $\tilde{N}^2 = (\tilde{\mathbf{W}}^2, \tilde{\boldsymbol{\theta}^2})$) le réseau

construit à partir de $\tilde{E}_{PG}$, comme suit : les lignes $2 < i \leq n$ de $N^1$ et $\tilde{N}^1$ (resp. de $N^2$ et $\tilde{N}^2$) sont identiques ; si la ligne 1 de $N^1$ (resp. $N^2$) est obtenue à partir du vecteur $\mathbf{V}^{i_1}$ de $PG_1$, alors la ligne 1 de $\tilde{N}^1$ (resp. $\tilde{N}^2$) est obtenue à partir du vecteur $\tilde{\mathbf{V}}^{i_1}$ de $\tilde{PG}_1$ ; si la ligne 2 de $N^1$ (resp. $N^2$) est obtenue à partir du vecteur $\mathbf{V}^{i_2}$ de $PG_1$, alors la ligne 2 de $\tilde{N}^1$ (resp. $\tilde{N}^2$) est obtenue à partir du vecteur $\tilde{\mathbf{V}}^{i_2}$ de $\tilde{PG}_1$.
Si l'on montre que $N^1 \tilde{\mathcal{R}} N^2 \Leftrightarrow \tilde{N}^1 \tilde{\mathcal{R}} \tilde{N}^2$, alors on a bien $N_{equiv} = \tilde{N}_{equiv}$, puisqu'à tout couple de réseaux équivalents construits à partir de $E_{PG}$, correspond un couple de réseaux équivalents construits à partir de $\tilde{E}_{PG}$, et inversement.
On suppose que $N^1 \tilde{\mathcal{R}} N^2$. D'après la proposition 2.4, il existe donc une permutation $\sigma \in S_n$ telle que $\mathbf{W}^2 = \sigma.\mathbf{W}^1$ et $\boldsymbol{\theta}^2 = \sigma.\boldsymbol{\theta}^1$, c'est-à-dire que chaque ligne $i \in \{1, \cdots, n\}$ de $N^1$ est égale à la ligne $\sigma(i)$ de $N^2$. $N^1$ et $N^2$ sont supposés distincts, donc $\sigma \neq Id$. De plus, $\sigma$ est telle que $\sigma(1) = 2$ et $\sigma(2) = 1$. En effet, comme on a supposé que $PG_1 = PG_2$ et $PG_1 \neq PG_i$, $2 < i \leq n$, les lignes 1 et 2 ne peuvent être permutées qu'entre elles. On ne s'intéresse pas aux lignes $2 < i \leq n$ puisque, par construction de $\tilde{N}^1$ (resp. $\tilde{N}^2$), les lignes $i$ de $N^1$ et $\tilde{N}^1$ (resp. de $N^2$ et $\tilde{N}^2$) sont identiques. Les vecteurs seuils ne posent pas de problème non plus, car $PG_1 = PG_2$ et $\tilde{PG}_1 = \tilde{PG}_2$ et donc $[\boldsymbol{\theta}^1]_1 = [\boldsymbol{\theta}^1]_2 = [\boldsymbol{\theta}^2]_1 = [\boldsymbol{\theta}^2]_2 = \theta$ et $[\tilde{\boldsymbol{\theta}}^1]_1 = [\tilde{\boldsymbol{\theta}}^1]_2 = [\tilde{\boldsymbol{\theta}}^2]_1 = [\tilde{\boldsymbol{\theta}}^2]_2 = \tilde{\theta}$ (on rappelle que $[\mathbf{V}]_i$ désigne l'élément i du vecteur $\mathbf{V}$). On s'intéresse donc uniquement aux éléments des lignes 1 et 2 des matrices d'interaction. Si on note, $w_j^{i_1}$ et $w_j^{i_2}$, $j \in \{1, \cdots, n\}$ les éléments $j$ de la ligne 1 et de la ligne 2 de la matrice $\mathbf{W}^1$, alors on a, pour tout $j \in \{1, \cdots, n\}$ :

- $[\mathbf{W}^2]_{1j} = w^{i_2}_{\sigma^{-1}(j)}$ et $[\mathbf{W}^2]_{2j} = w^{i_1}_{\sigma^{-1}(j)}$, puisque $\mathbf{W}^2 = \sigma.\mathbf{W}^1$, avec $\sigma(1) = 2$ et $\sigma(2) = 1$ (on rappelle que $[\mathbf{M}]_{ij}$ désigne l'élément de la matrice $\mathbf{M}$ qui se trouve à la ligne $i$ et à la colonne $j$) ;
- $[\tilde{\mathbf{W}}^1]_{1j} = \phi(w_j^{i_1})$ et $[\tilde{\mathbf{W}}^1]_{2j} = \phi(w_j^{i_2})$, par construction de $\tilde{N}^1$ à partir de $N^1$ ;
- $[\tilde{\mathbf{W}}^2]_{1j} = \phi(w^{i_2}_{\sigma^{-1}(j)})$ et $[\tilde{\mathbf{W}}^2]_{2j} = \phi(w^{i_1}_{\sigma^{-1}(j)})$, par construction de $\tilde{N}^2$ à partir de $N^2$.

Comme pour tout $j \in \{1, \cdots, n\}$, on a $\tilde{w}_j^{i_1} = \phi(w_j^{i_1})$ et $\tilde{w}_j^{i_2} = \phi(w_j^{i_2})$, on peut écrire, pour tout $j \in \{1, \cdots, n\}$ :

- $[\tilde{\mathbf{W}}^1]_{1j} = \tilde{w}_j^{i_1}$ et $[\tilde{\mathbf{W}}^1]_{2j} = \tilde{w}_j^{i_2}$ ;
- $[\tilde{\mathbf{W}}^2]_{1j} = \tilde{w}^{i_2}_{\sigma^{-1}(j)}$ et $[\tilde{\mathbf{W}}^2]_{2j} = \tilde{w}^{i_1}_{\sigma^{-1}(j)}$.

On a donc $\tilde{\mathbf{W}}^2 = \sigma.\tilde{\mathbf{W}}^1$, d'où $\tilde{N}^1 \tilde{\mathcal{R}} \tilde{N}^2$.
On montre l'autre sens de l'équivalence de la même façon, en utilisant $\phi^{-1}$ à la place de $\phi$. □

*Exemple* 2.12 *:* Soit $E_{PG} = \{PG_1, PG_2, PG_3\}$ l'ensemble des 3 parties génératrices d'éléments de $FS_3/\mathcal{R}$ telles que $PG_1 = PG_2$, et dont les vecteurs représentatifs sont $\mathbf{V}^1 = (-1, -1, -2)(-3)$ (représentatif de $PG_1 = PG_2$) et $\mathbf{V}^3 = (0, 0, 0)(0)$ (représentatif de $PG_3$).
Soit $\tilde{E}_{PG} = \{\tilde{PG}_1, \tilde{PG}_2, \tilde{PG}_3\}$ l'ensemble des 3 parties génératrices d'éléments de $FS_3/\mathcal{R}$ telles que $\tilde{PG}_1 = \tilde{PG}_2$, et dont les vecteurs représentatifs sont $\tilde{\mathbf{V}}^1 = (-1, 0, 0)(-1)$ (représentatif de $\tilde{PG}_1 = \tilde{PG}_2$) et $\tilde{\mathbf{V}}^3 = (1, 1, 1)(0)$ (représentatif

de $\tilde{PG}_3$).
On voit que $\mathbf{V}^1$ et $\tilde{\mathbf{V}}^1$ sont de structure identique, tout comme $\mathbf{V}^3$ et $\tilde{\mathbf{V}}^3$. $E_{PG}$ et $\tilde{E}_{PG}$ sont donc de structure générale identique.
D'après la proposition précédente, les nombres de classes d'équivalence de $RS_3/\tilde{\mathcal{R}}$ que l'on peut générer à partir de $E_{PG}$ et $\tilde{E}_{PG}$ sont donc égaux. On peut majorer ce nombre par $3 \times 3 \times 1 = 9$, puisque $PG_1$ et $PG_2$ contiennent 3 vecteurs d'interaction (les 3 vecteurs d'interaction obtenus en permutant les poids de $\mathbf{V}^1$), alors que $PG_3$ contient uniquement $\mathbf{V}^3$.

Pour calculer le nombre total de classes d'équivalence de RABS à $n$ éléments, on peut donc regrouper les ensembles de parties génératrices de structure générale identique. La théorie des actions de groupe nous permet de déterminer le nombre exact de classes d'équivalence définies à partir d'un ensemble $E_{PG}$ de $n$ parties génératrices d'éléments de $FS_n/\tilde{\mathcal{R}}$.

**Définition 2.15** (Action de groupe)
*Soit $G$ un groupe dont la loi est notée multiplicativement.*
*On peut définir une* action du groupe *$G$ sur un ensemble $E$ par une application :*
$G \times E \to E$
$(g, x) \mapsto g.x$
*qui vérifie* $\forall g, g' \in G,\ \forall x \in E,\ \ g'.(g.x) = (g'.g).x$

Soit $E_{RC}$ l'ensemble des réseaux construits à partir de $E_{PG}$. L'application $\Phi$ : $\begin{cases} S_n \times E_{RC} \to E_{RC} \\ (\sigma, (\mathbf{W}, \boldsymbol{\theta})) \mapsto (\sigma.\mathbf{W}, \sigma.\boldsymbol{\theta}) \end{cases}$ définit une action du groupe symétrique sur $E_{RC}$ (la loi de $S_n$ est la composition de permutations). Par définition de l'ensemble $E_{RC}$, si $(\mathbf{W}, \boldsymbol{\theta}) \in E_{RC}$, alors $\forall \sigma \in S_n, (\sigma.\mathbf{W}, \sigma.\boldsymbol{\theta}) \in E_{RC}$. De plus, on a bien $\forall \sigma, \mu \in S_n$,

$$\begin{aligned} (\mu.\sigma).\mathbf{W} &= \mathbf{P}_{\mu.\sigma}\mathbf{W}\mathbf{P}_{\mu.\sigma}^{-1} \\ &= (\mathbf{P}_\mu\mathbf{P}_\sigma).\mathbf{W}.(\mathbf{P}_\mu\mathbf{P}_\sigma)^{-1} \\ &= \mathbf{P}_\mu\mathbf{P}_\sigma.\mathbf{W}.\mathbf{P}_\sigma^{-1}\mathbf{P}_\mu^{-1} \\ &= \mathbf{P}_\mu.(\sigma.\mathbf{W}).\mathbf{P}_\mu^{-1} \\ &= \mu.(\sigma.\mathbf{W}) \end{aligned}$$

De même $(\mu.\sigma).\boldsymbol{\theta} = \mu.(\sigma.\boldsymbol{\theta})$.

**Définition 2.16** (Orbite et éléments fixés)
*Soit $x$ un élément de $E$.*
*L'*orbite *de $x$, notée $O_x$, est définie par : $O_x = \{g.x,\ g \in G\}$.*

*Soit $g$ un élément de $G$.*
*L'ensemble des* éléments fixés *par $g$, noté $Fix_g$, est défini par :*
$Fix_g = \{x \in E\ ;\ g.x = x\}$.

Dans notre cas, par définition de $\Phi$, l'orbite d'un élément $(\mathbf{W}, \boldsymbol{\theta}) \in E_{RC}$ est l'ensemble des éléments de $RS_n$ équivalents à $(\mathbf{W}, \boldsymbol{\theta})$ selon $\tilde{\mathcal{R}}$. Chaque orbite représente donc une classe d'équivalence de $RS_n/\tilde{\mathcal{R}}$. La formule de Burnside nous permet de dénombrer ces classes d'équivalence.

**Théorème 2.1** (Formule de Burnside)
*On désigne par $\Omega$ l'ensemble des orbites, $\Omega = \{O_x,\ x \in E\}$.*
*Si $G$ et $E$ sont finis, alors le nombre d'orbites est*

$$\#\Omega = \frac{1}{\#G} \sum_{g \in G} \#Fix_g.$$

On a donc la proposition immédiate suivante.

**Proposition 2.6**
*Soit $E_{PG} = \{PG_1, \cdots, PG_n\} \subset Gen(FS_n/\mathcal{R})^n$ un ensemble de n parties génératrices d'éléments de $FS_n/\mathcal{R}$. On note $n_i$ la taille de l'élément $PG_i$, $i \in \{1, \cdots, n\}$.*

*Si les éléments de $E_{PG}$ sont 2 à 2 distincts, alors $N_{equiv}$, le nombre de classes d'équivalence de $RS_n/\tilde{\mathcal{R}}$ définies par les réseaux construits à partir de $E_{PG}$ est :*

$$N_{equiv} = \prod_{i=1}^{n} n_i.$$

*Preuve.* On suppose que les éléments de $E_{PG}$ sont 2 à 2 distincts. On note $E_{RC}$ l'ensemble des réseaux construits à partir de $E_{PG}$. La formule de Burnside s'écrit alors :

$$N_{equiv} = \frac{1}{n!} \sum_{\sigma \in S_n} \#\{(\mathbf{W}^e, \boldsymbol{\theta}^e) \in E_{RC}\ ;\ (\mathbf{W}^e, \boldsymbol{\theta}^e) = (\sigma.\mathbf{W}^e, \sigma.\boldsymbol{\theta}^e)\}$$

Pour tout $\sigma \in S_n$, pour tout $i, j \in \{1, \cdots, n\}$, on a $[\sigma.\mathbf{W}^e]_{\sigma(i)\,\sigma(j)} = [\mathbf{W}^e]_{ij}$. Comme les élements de $E_{PG}$ sont 2 à 2 distincts, les lignes $i$ et $\sigma(i)$, $i \in \{1, \cdots, n\}$, correspondent 2 à 2, uniquement si $\sigma = Id$, c'est-à-dire $\sigma.\mathbf{W}^e = \mathbf{W}^e \Leftrightarrow \sigma = Id$. Dans le cas où les éléments de $E_{PG}$ sont 2 à 2 distincts, on a donc :

$$\begin{aligned} N_{equiv} &= \frac{1}{n!}\#\{(\mathbf{W}^e, \boldsymbol{\theta}^e) \in E_{RC}\ ;\ (\mathbf{W}^e, \boldsymbol{\theta}^e) = (\mathbf{W}^e, \boldsymbol{\theta}^e)\} \\ &= \frac{1}{n!}\#E_{RC} \\ &= \frac{1}{n!} n! \prod_{i=1}^{n} n_i \\ &= \prod_{i=1}^{n} n_i \end{aligned}$$

□

Pour les ensembles de parties génératrices à $n$ éléments distincts, le nombre de classes d'équivalence de RABS que l'on peut construire est donc donné de façon théorique. Pour les ensembles dans lesquels il y a au moins 2 parties génératrices égales, le nombre exact de classes d'équivalence est déterminé par le calcul, en utilisant la formule de Burnside adaptée à notre problème, et on connaît également un majorant de ce nombre.

*Exemple* 2.13 *:* On reprend l'ensemble de 3 parties génératrices d'éléments de $FS_3/\mathcal{R}$ noté $E_{PG}$ de l'exemple 2.12. Nous avions vu que le nombre de classes d'équivalence de $\#RS_3/\tilde{\mathcal{R}}$ générées à partir de $E_{PG}$ est majoré par 9. D'après le calcul fait à partir de la formule de Burnside, on peut générer exactement 6 classes d'équivalence de RABS à 3 éléments à partir de $E_{PG}$. On remarquera que la formule de Burnside n'est pas constructive : elle nous permet de calculer le nombre de classes d'équivalence, mais on ne peut pas en déduire un représentant pour chacune de ces classes.

Pour la suite du manuscrit, on considère que la population totale de réseaux ne contient que des RABS non équivalents selon $\tilde{\mathcal{R}}$, c'est-à-dire $\#RS_n/\tilde{\mathcal{R}}$ éléments. On assimile donc tout réseau à sa classe d'équivalence. Nous avons en effet déjà constaté que 2 RABS équivalents sont en fait un seul et même réseau dont on a permuté les éléments. On considère donc que la formule de Burnside nous donne le nombre total de réseaux différents que l'on peut construire à partir d'un ensemble de parties génératrices à $n$ éléments.

Sans faire de calcul exact à partir de la formule de Burnside, on peut donner un minorant et un majorant du nombre de RABS différents à $n$ éléments. Ils sont obtenus à partir de deux populations de RABS à $n$ éléments, qui encadrent, au sens de l'inclusion, la population de tous les RABS différents à $n$ éléments. Le première population, que nous appellerons population minorante, contient tous les RABS (non équivalents) construits à partir des sélections (ou ensembles) de $n$ éléments 2 à 2 distincts de $Gen(FS_n/\mathcal{R})$. Cette population ne contient pas tous les RABS différents à $n$ éléments, puisqu'elle ne contient pas ceux construits à partir de sélections de $n$ éléments de $Gen(FS_n/\mathcal{R})$, dont au moins 2 sont égaux. Le nombre de RABS de cette population est un minorant du nombre total de RABS différents à $n$ éléments. On note ce minorant, $Mino_n$. La deuxième population, que nous appellerons population majorante, contient tous les RABS à $n$ éléments que l'on peut construire à partir de toutes les sélections de $n$ éléments de $Gen(FS_n/\mathcal{R})$. Dans cette population, on ne fait pas la distinction entre les sélections de $n$ éléments distincts de $Gen(FS_n/\mathcal{R})$ et les sélections pour lesquelles au moins 2 éléments sont égaux. Ainsi, pour chaque sélection $E_{PG} = (PG_1, \cdots, PG_n)$, on conserve les $\prod_{i=1}^{n} n_i$ RABS construits ($n_i$ représente le nombre de vecteurs d'interaction que la partie génératrice $PG_i$ contient, pour $i \in \{1, \cdots, n\}$). Cette population contient donc tous les RABS non équivalents à $n$ éléments, plus certains RABS équivalents aux premiers. Le nombre de RABS de cette population est un majorant du nombre total de RABS différents à $n$ éléments. On note ce majorant, $Majo_n$.

Pour $1 \leq n \leq 5$, nous avons calculé le nombre exact de RABS différents, $\#RS_n/\tilde{\mathcal{R}}$, à partir de la formule de Burnside. Le tableau 2.5 récapitule, pour $1 \leq n \leq 7$, les valeurs de $Mino_n$, $Majo_n$, et de $\#RS_n/\tilde{\mathcal{R}}$ pour $1 \leq n \leq 5$ uniquement. La différence $Diff_n = Majo_n - Mino_n$, exprimée en pourcentage de $Mino_n$ est également donnée.

La différence entre le majorant et le minorant du nombre exact de RABS à $n$

TAB. 2.5 – Minorant ($Mino_n$), valeur exacte ($\#RS_n/\tilde{\mathcal{R}}$) et majorant ($Majo_n$) du nombre de réseaux d'automates booléens à seuil à $1 \leq n \leq 7$ éléments ; $Diff_n$ : différence entre le majorant et le minorant, exprimée en pourcentage du minorant.

| $n$ | $Mino_n$ | $\#RS_n/\tilde{\mathcal{R}}$ | $Majo_n$ | $Diff_n$ |
|---|---|---|---|---|
| 1 | 4 | 4 | 4 | $0,00\%$ |
| 2 | 87 | 105 | 109 | $25,29\%$ |
| 3 | 167 084 | 188 968 | 209 100 | $25,15\%$ |
| 4 | $4,99 \times 10^{11}$ | $5,23 \times 10^{11}$ | $5,47 \times 10^{11}$ | $9,72\%$ |
| 5 | $6,25 \times 10^{22}$ | $6,30 \times 10^{22}$ | $6,36 \times 10^{22}$ | $1,64\%$ |
| 6 | $1,60 \times 10^{40}$ | - | $1,60 \times 10^{40}$ | $0,10\%$ |
| 7 | $3,02 \times 10^{56}$ | - | $3,02 \times 10^{56}$ | $0,04\%$ |

éléments diminue avec $n$ (si l'on excepte le cas $n = 1$). Ceci s'explique par le fait que, plus $n$ est grand, plus le nombre de sélections de $n$ éléments différents parmi $\#FS_n/\mathcal{R}$ se rapproche du nombre de sélections de $n$ éléments parmi $\#FS_n/\mathcal{R}$. Autrement dit, plus $n$ est grand, moins il est probable d'avoir au moins 2 éléments égaux dans une sélection de $n$ éléments choisis parmi $\#FS_n/\mathcal{R}$. Par exemple, pour $n = 6$, nous avons vu que le nombre de classes d'équivalence des fonctions booléennes à seuil est $\#FS_6/\mathcal{R} = 38\,456$, il y a donc $4,494 \times 10^{24}$ sélections de 6 éléments quelconques possibles contre $4,490 \times 10^{24}$ sélections de 6 éléments différents. Ainsi, prendre le nombre de RABS de la population minorante ou celui de la population majorante ne change presque rien pour $n$ suffisamment grand. On décide qu'à partir de $n = 6$, la différence est infime, et que l'on peut alors prendre le majorant comme nombre de RABS à $n$ éléments (on considère que la population de RABS à $n$ éléments et la population majorante sont identiques). Ceci nous permet donc de ne pas faire le calcul à partir de la formule de Burnside pour ces 2 tailles.

On voit aussi, d'après la tableau 2.5, que l'on peut simuler la dynamique de tous les réseaux différents, pour les tailles 1, 2 et 3. En revanche, à partir de $n = 4$, on ne peut plus simuler la dynamique de la population totale de RABS, car ceux-ci sont trop nombreux. Pour les tailles supérieures à 3, nous simulons donc un échantillon de la population totale de réseaux. Les algorithmes pour réaliser ces simulations sont présentés dans la partie suivante.

Avant de clore cette partie, nous allons aborder un point important concernant la population de réseaux générés. Nous avons vu dans la partie précédente (tableau 2.2) que, pour $n \geq 4$, la proportion de fonctions booléennes à seuil dont le nombre d'arguments exact est inférieur strictement à $n$ parmi les fonctions booléennes à seuil à $n$ arguments était très faible. Ainsi, lorsque l'on représente les fonctions booléennes à seuil à $n$ arguments par des vecteurs d'interaction, très peu de vecteurs ont des poids d'interaction nuls. En construisant les réseaux à partir de ces vecteurs d'interaction, on a donc très peu de poids nuls dans les

matrices d'interaction. Autrement dit, la connectivité moyenne des graphes d'interaction de RABS à $n$ éléments est très élevée, comme le montre le tableau 2.6. Nous avions annoncé dans la partie 1.2.2, que les graphes d'interaction de réseaux de régulation génétique connus avaient une connectivité comprise entre $\frac{3}{2}$ et 3. Les RABS à $n$ éléments ont des graphes dont la connectivité est de l'ordre de $n$. Ce sont quasiment des réseaux à graphe d'interaction complet. Nous verrons dans le chapitre suivant que la connectivité des graphes d'interaction et la robustesse des RABS sont liées. Nous n'abordons pas plus en détail l'indice moyen d'inhibition des graphes d'interaction, car nous avons vu dans la partie précédente que le nombre de fonctions à seuil à exactement $n$ arguments avec $k \leq n$ activations est égal au nombre de fonctions à seuil à exactement $n$ arguments avec $k \leq n$ inhibitions. Ainsi, on peut considérer que l'indice moyen d'inhibition des graphes d'interaction de RABS à $n$ éléments est égal à $\frac{1}{2}$ quelle que soit la valeur de $n$, ce qui correspond aux indices moyens d'inhibition observés pour les réseaux de régulation génétique.

TAB. 2.6 – Connectivité moyenne des graphes d'interaction de réseaux d'automate booléen à seuil à $n$ éléments, notée $CM_n$.

| $n$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $CM_n$ | 0,5 | 1,43 | 2,60 | 3,78 | 4,90 | 5,97 | 7,00 |

Dans cette partie, nous avons exposé comment on obtient tous les réseaux booléens à seuil à $n$ éléments à partir des classes d'équivalence des fonctions booléennes à seuil à $n$ arguments. Nous avons donné le nombre de ces réseaux pour $1 \leq n \leq 7$. D'après ces nombres, nous avons vu qu'il était impossible de simuler la dynamique de tous les RABS à plus de 3 éléments. Dans la partie suivante, nous allons donner l'algorithme qui nous permet de simuler la dynamique de tous les RABS à 3 éléments ou moins, et d'un échantillon de la population totale des RABS à $n$ éléments pour $4 \leq n \leq 7$.

## 2.4 Algorithme pour la simulation des réseaux

Dans cette partie, nous allons présenter l'algorithme qui nous permet de simuler le comportement dynamique des réseaux d'automates booléens à seuil. C'est à partir de cet algorithme que sont réalisées les simulations correspondant aux résultats du chapitre 3. Nous avons vu dans la partie précédente que, selon le nombre d'éléments des RABS dont on veut simuler la dynamique, nous pouvons faire des simulations sur la population entière des RABS, ou sur un échantillon de celle-ci uniquement. Dans la suite, nous parlerons d'une part de simulation exhaustive de la dynamique des petits réseaux (réseaux ayant au plus 3 éléments), d'autre part, de simulation de la dynamique d'un échantillon de grands réseaux (réseaux ayant au moins 4 éléments). On voit que le terme "grands" est ici tout

relatif. Que ce soit pour les petits réseaux ou pour les grands réseaux, notre algorithme de simulation se décompose en deux étapes distinctes. La première phase consiste à générer les réseaux. Il en existe donc deux variantes différentes, selon la taille des réseaux que l'on veut simuler. La deuxième phase est la phase de la simulation du comportement dynamique des réseaux générés. Cette phase est la même quelle que soit la taille des réseaux concernés. Nous avons vu dans la partie 1.3.2, que l'on peut définir 4 classes de comportement dynamique pour les réseaux d'automates. A partir des simulations, nous pourrons donc estimer la proportion de réseaux correspondant à chaque classe de comportement au sein de la population totale des RABS. Nous expliquons dans cette partie comment sont estimées ces proportions. Avant cela, nous présentons tout d'abord la phase de simulation de la dynamique d'un réseau, puis, nous abordons l'algorithme de génération des réseaux.

### Simulation de la dynamique

Le but de nos simulations est de décrire le comportement dynamique des RABS. Comme nous l'avons présenté dans la partie 1.3.2, nous nous intéressons aux conséquences d'un changement dans le mode d'itération blocs-séquentiel avec lequel le réseau est itéré, sur les attracteurs de ce réseau. Afin d'étudier les attracteurs d'un réseau, il nous faut donc simuler l'ensemble des trajectoires correspondant à toutes les configurations initiales possibles du réseau. Une configuration d'un RABS à $n$ éléments est un $n$-uplet de valeurs booléennes. Il y a donc $2^n$ configurations possibles pour un RABS à $n$ éléments. Si l'on connaît, pour chacune des configurations du réseau, la configuration suivante, obtenue en appliquant la fonction de transition globale du réseau, alors on sait décrire toutes les trajectoires. Ceci est la conséquence du déterminisme de la fonction de transition globale des RABS. Ainsi, à partir de chaque configuration du réseau à l'instant $t$, on obtient une unique configuration image par la fonction de transition globale, qui est la configuration du réseau à l'instant $t + 1$. Pour un mode d'itération blocs-séquentiel donné, on connaît donc l'ensemble des attracteurs atteints par le réseau étudié, en calculant les $2^n$ transitions entre configurations de ce réseau. La procédure 1 présente l'algorithme utilisé, écrit en pseudo-code de programmation.

A l'issue de cette procédure, on a l'ensemble des points fixes et des cycles limites que le réseau atteint, pour tous les modes d'itération blocs-séquentiels. On a donc une description totale du comportement dynamique du réseau. On peut en étudier la robustesse face à un changement de mode d'itération, par exemple.

### Génération des réseaux

Nous avons déjà expliqué que, selon le nombre d'éléments que contiennent les réseaux étudiés, on pouvait simuler le comportement dynamique pour la population entière ou bien se restreindre à l'étude d'un échantillon de cette population.

**Procédure 1** Etape de simulation de la dynamique.

```
1: for (m ∈ {modes d'itération}) do
2:   PF(m) ← ∅        // points fixes pour le mode m
3:   CL(m) ← ∅        // cycles limites pour le mode m
4:   A_traiter← {toutes configurations}        // configurations à traiter
5:   Traitées← ∅        // configurations traitées
6:   for (i ∈ A_traiter) do
7:     Fin_traject ← FAUX
8:     vider(TC)        // on vide la trajectoire courante
9:     i_TC ← 1
10:    TC(i_TC) ← i        // trajectoire courante
11:    i_TC++
12:    while (Fin_traject=FAUX) do
13:      A_traiter ← A_traiter \ i        // on supprime i de A_traiter
14:      j ← config_suivante(i)        // calcul de la transition
15:      if (j = i) then
16:        PF(m) ← PF(m) ∪ {i}
17:        Traitées ← Traitées ∪ TC
18:        Fin_traject ← VRAI
19:      else if (∃k/ TC(k) = j) then
20:        CL(m) ← CL(m) ∪ TC(k ::taille(TC))
21:        Traitées ← Traitées ∪ TC
22:        Fin_traject ← VRAI
23:      else if (j ∈ Traitées) then
24:        Traitées ← Traitées ∪ TC
25:        Fin_traject ← VRAI
26:      else
27:        TC(i_TC) ← j
28:        i_TC++
29:        i ← j
30:      end if
31:    end while
32:  end for
33: end for
```

Pour un nombre d'éléments $n$ inférieur strictement à 4, nous construisons tous les RABS différents (non équivalents selon $\tilde{\mathcal{R}}$). Pour cela, on utilise la procédure décrite dans la partie 2.3. On suppose que l'on a à disposition l'ensemble de tous les vecteurs d'interaction représentatifs des classes d'équivalence des fonction booléennes à seuil à $n$ arguments. A partir de chacun de ces vecteurs d'interaction représentatifs, on peut construire la partie génératrice de la classe d'équivalence associée au vecteur, comme expliqué dans la définition 2.12. On peut donc générer toutes les sélections de $n$ parties génératrices. On note $E_{PG}$ une sélection de $n$ parties génératrices et $n_i$, $i \in \{1, \cdots, n\}$ le nombre de vecteurs d'interaction de l'élément $i$ de $E_{PG}$. Comme nous l'avons expliqué dans la partie 2.3, on peut construire $\prod_{i=1}^{n} n_i$ RABS à $n$ élément à partir de $E_{PG}$. Si les parties génératrices de $E_{PG}$ sont 2 à 2 distinctes, alors, les $\prod_{i=1}^{n} n_i$ réseaux sont tous différents, d'après la proposition 2.6. Si parmi les $n$ parties génératrices de $E_{PG}$, au moins 2 sont égales, alors, on supprime tous les doublons de réseaux équivalents $((\mathbf{W}, \boldsymbol{\theta}), (\mathbf{W}', \boldsymbol{\theta}'))$, en testant, pour chaque permutation $\sigma \in S_n$, si $\mathbf{W}' = \sigma.\mathbf{W}$ et $\boldsymbol{\theta}' = \boldsymbol{\theta}$. En construisant ainsi tous les réseaux différents à partir de toutes les sélections de $n$ parties génératrices, on construit la population totale de RABS à $n$ éléments.

Pour un nombre d'éléments $n$ supérieur ou égal à 4, nous construisons un échantillon de la population totale de RABS à $n$ éléments. Evidemment, nous souhaitons que cet échantillon soit représentatif de la population totale, et donc, que tous les RABS à $n$ éléments aient la même probabilité d'être sélectionnés de façon indépendante lors de la construction de l'échantillon. On note $N_{sample}$ la taille de l'échantillon et $N_n$ le nombre de RABS différents à $n$ éléments ($N_n = \#RS_n/\tilde{\mathcal{R}}$ pour $4 \leq n \leq 5$ et $N_n = Majo_n$ pour $6 \leq n \leq 7$). A chaque sélection d'un élément de l'échantillon, on note $p_r$ la probabilité pour que le réseau $r$ soit sélectionné. Quel que soit $r$, on veut $p_r = p = \frac{1}{N_n}$. Les éléments de l'échantillon sont sélectionnés de façon indépendante.

Les réseaux sont toujours obtenus à partir des sélections de $n$ parties génératrices de classes d'équivalence de fonction booléennes à seuil. Chacun des RABS à $n$ éléments ne peut être construit qu'à partir d'une unique sélection de $n$ parties génératrices. On note $E_{PG}$ une sélection de $n$ parties génératrices et $n_i$, $i \in \{1, \cdots, n\}$ le nombre de vecteurs d'interaction de l'élément $i$ de $E_{PG}$. Comme nous l'avons expliqué dans la partie 2.3, on peut construire $\prod_{i=1}^{n} n_i$ RABS à $n$ éléments à partir de $E_{PG}$. Ces réseaux ne sont pas forcément tous non équivalents 2 à 2, or on ne s'intéresse qu'aux réseaux non équivalents entre eux. On fait ici l'hypothèse simplificatrice suivante : en tirant de façon uniforme des réseaux dans l'ensemble de tous les réseaux construits à partir de $E_{PG}$, la répartition des réseaux non équivalents entre eux obtenus est uniforme. Cette hypothèse est valide dans le cas où tous les éléments de $E_{PG}$ sont distincts, et pour $n \geq 6$, puisque nous travaillons alors avec la population majorante. Si $N_{E_{PG}}$ est le nombre de classes d'équivalence de $RS_n/\tilde{\mathcal{R}}$ obtenues à partir de $E_{PG}$, chaque réseau de la population des RABS non équivalents entre eux, obtenu à partir de $E_{PG}$, a une

probabilité égale à $\frac{1}{N_{E_{PG}}}$ d'être tiré au hasard parmi tous les réseaux construits à partir de $E_{PG}$.

Pour chacun des $N_{sample}$ éléments de l'échantillon, nous tirons au hasard 1 sélection de $n$ parties génératrices. A partir de cette sélection, nous tirons au hasard un réseau de la population des réseaux non équivalents entre eux, comme expliqué dans le paragraphe précédent. Ainsi, pour tout RABS à $n$ éléments, noté $r$, si l'on note $E^r_{PG}$ la sélection de $n$ parties génératrices à partir de laquelle $r$ est construit, et A, l'événement "le réseau $r$ est sélectionné", B, l'événement "le réseau $r$ est tiré au hasard parmi les réseaux construits à partir de $E^r_{PG}$", et enfin, C, l'événement "la sélection $E^r_{PG}$ est sélectionnée", alors, on a : $p = P(A) = P(B|C)P(C)$. Nous avons vu dans le paragraphe précédent que l'on suppose que $P(B|C) = \frac{1}{N_{E^r_{PG}}}$, avec $N_{E^r_{PG}}$, le nombre de classes d'équivalence de $RS_n/\tilde{\mathcal{R}}$ obtenues à partir de $E^r_{PG}$. Afin d'avoir $p = \frac{1}{N_n}$, il faut donc $P(C) = \frac{N_{E^r_{PG}}}{N_n}$. Le tirage au hasard des sélections de $n$ parties génératrices doit donc se faire de telle sorte que chaque sélection est tirée avec une probabilité proportionnelle au nombre de réseaux non équivalents entre eux que l'on peut construire à partir de cette sélection. Comme nous l'avons vu dans la partie 2.3, ces nombres de réseaux sont donnés par la formule de Burnside.

Si l'on imagine que l'ensemble des RABS à $n$ éléments, $RS_n/\tilde{\mathcal{R}}$, est représenté par un carré dont l'aire est égale à $N_n$, on peut paver ce carré à l'aide de formes géométriques planes élémentaires, d'aire 1. Chacune de ces formes élémentaires représente un réseau de la population. On peut colorier ces formes de façon à ce que l'ensemble des formes de la même couleur représente l'ensemble des réseaux (non équivalents) générés à partir d'une même sélection de $n$ parties génératrices d'éléments de $FS_n/\mathcal{R}$ (on suppose que l'on a une infinité de couleurs à disposition). Construire un échantillon de $N_{sample}$ RABS à $n$ éléments, revient à prendre de façon uniforme sur la surface, $N_{sample}$ points dans le carré. Les formes élémentaires dans lesquelles "tombent" les points sont sélectionnées. Pour chaque point pris dans le carré, deux formes élémentaires ont la même probabilité d'être sélectionnées, $\frac{1}{N_n}$. Pour chaque couleur $c$, la probabilité qu'une des formes élémentaires de couleur $c$ soit ainsi sélectionnée, est égale au nombre de formes élémentaires de la couleur $c$ sur le nombre total de formes élémentaires, $N_n$.

Nous avons donc une procédure de construction d'échantillons de $N_{sample}$ RABS à $n$ éléments, qui assure que nos échantillons sont représentatifs de la population de RABS non équivalents entre eux. A partir de ces échantillons de RABS, on peut réaliser des simulations sur le comportement dynamique des RABS.

### Estimation de proportions

Dans le chapitre 3, sont exposés les résultats des simulations menées sur les RABS à $n$ éléments, grâce aux procédures présentées dans les paragraphes précédents. Nous étudions notamment la proportion de réseaux dont le comportement dynamique correspond à chacune des classes de comportement définies dans la

partie 1.3.2. Pour les petits réseaux, nous pouvons simuler la population totale de réseaux, ainsi la proportion observée à partir des simulations pour chaque classe de comportement correspond effectivement à la proportion dans la population totale. Dans le cas des grands réseaux, les simulations se font sur des échantillons de la population totale.

Nous observons donc la proportion de chaque classe de comportement dans l'échantillon. Afin d'estimer la proportion dans la population totale à partir de cette proportion observée, nous utilisons l'estimateur classique : l'estimation de la proportion dans la population totale est la proportion observée pour l'échantillon. A cette estimation ponctuelle, on peut associer une estimation sous la forme d'un intervalle qui représente la confiance que l'on peut accorder à l'estimation. Cette confiance dépend notamment de la taille de l'échantillon. L'intervalle s'appelle intervalle de confiance et est centré en l'estimation de la proportion. On peut théoriquement donner un intervalle de confiance pour n'importe quel niveau de signification $1-\alpha$, c'est-à-dire un intervalle pour lequel on considère qu'il y a une probabilité égale à $\alpha$ que la proportion dans la population ne soit pas comprise dans l'intervalle de confiance. En général, on donne des intervalles de confiance pour un niveau de signification égal à 95%. Ainsi, la probabilité pour que la proportion dans la population ne soit pas comprise dans l'intervalle donné n'est que de 5%. Si l'on note $p$ la proportion observée pour un échantillon de taille $N_{sample}$, alors l'intervalle de confiance de niveau de signification 95% pour la proportion dans la population est donné par :

$$\left[ p - \sqrt{\frac{p(1-p)}{N_{sample}}} u_{5\%} \; ; \; p + \sqrt{\frac{p(1-p)}{N_{sample}}} u_{5\%} \right]$$

avec $u_{5\%} = 1.960$. Nous présenterons les intervalles de confiance (de niveau 95%) sous la forme $p \pm \delta_p$ plutôt que $[p - \delta_p; p + \delta_p]$.

Remarquons enfin que pour être tout à fait exact, dans le cas $n \geq 6$, il faudrait appliquer un coefficient correcteur aux bornes des intervalles de confiance. En effet, si l'on tient compte du fait que l'échantillon de réseaux simulés est construit à partir de la population majorante, il faut corriger l'estimation pour que le niveau de signification soit exact pour la population qui contient une seule fois tous les RABS non équivalents à $n$ éléments. Etant donnée l'infime différence entre les populations minorante et majorante (et donc la différence encore plus faible entre les populations majorante et exacte) pour $n \geq 6$, nous n'appliquons pas ce coefficient correcteur.

Dans ce chapitre, nous avons abordé un point qui nous semble important dans la modélisation des réseaux de régulation génétique par des réseaux d'automates booléens à seuil : la représentation des fonctions booléennes à seuil par des vecteurs d'interaction que l'on a appelés vecteurs minimaux. Cette représentation nous assure que chaque poids non nul de la matrice d'interaction d'un RABS modélise effectivement soit une inhibition, soit une activation dans le réseau de

régulation. A partir des fonctions booléennes à seuil ainsi définies, nous avons montré comment on peut générer l'ensemble de tous les RABS différents pour un nombre d'éléments donné. Enfin, nous avons décrit les procédures qui nous permettent de faire des simulations concernant la dynamique des RABS. Dans le chapitre suivant, nous allons présenter les résultats issus de ces simulations.

# Chapitre 3

# Résultats expérimentaux

Notre travail vise à mettre en exergue la robustesse des réseaux de régulation face à un changement de mode d'itération. Nous modélisons les réseaux de régulation par des réseaux d'automates booléens à seuil. Nous avons vu, dans le chapitre précédent, que le nombre de réseaux d'automates booléens à seuil à $n$ éléments, pour $n \geq 4$, est bien trop important pour espérer faire une étude exhaustive de la dynamique de ces réseaux sur l'ensemble de la population. Dans ce chapitre, nous allons donc présenter des résultats de simulations effectuées sur la population totale de RABS de taille inférieure à 4, et sur des échantillons de population des RABS à $n$ éléments, pour $4 \leq n \leq 7$. Ces simulations concernent le caractère dynamique des réseaux. Tout comme lors de la construction des vecteurs d'interaction représentatifs des classes d'équivalence des fonctions booléennes à seuil, les simulations ont été en grande partie effectuées sur le calculateur Healthphy (CIMENT-Grenoble). L'algorithme présenté dans la partie 2.4, a été implémenté sous la forme d'un programme parallèle, exécuté sur 16 à 20 processeurs simultanément.

Dans une première partie, nous donnerons des statistiques générales obtenues concernant le comportement dynamique des RABS à $1 \leq n \leq 7$ éléments. Ces statistiques montrent notamment qu'en moyenne, la robustesse des réseaux face à un changement de mode d'itération blocs-séquentiel diminue, lorsque la taille des réseaux augmente. Pour cette raison, nous étudions de façon plus approfondie la classe de comportement dynamique *Ev* décrite dans la partie 1.3.2. Cette classe est celle des réseaux dont le comportement dynamique change le plus avec les modes d'itération. Nous traitons ensuite un exemple de modèle de réseau de régulation génétique, celui de la morphogénèse florale chez *Arabidopsis thaliana*. Nous montrons que l'on peut décrire de façon exhaustive le comportement dynamique de ce réseau de 12 éléments, en se ramenant à l'étude de 2 réseaux de taille 3. Enfin, nous proposons une première approche d'étude de l'influence des $\mu$ARNs sur la robustesse des réseaux de régulation génétique.

## 3.1 Statistiques générales

Dans cette partie, nous allons donner des statistiques générales obtenues par simulation de la dynamique des RABS à $n$ éléments, avec $1 \leq n \leq 7$. Nous étudions tout d'abord la répartition des RABS dans les 4 classes de comportement définies dans la partie 1.3.2, en fonction de la taille des réseaux. Cette répartition nous permet de donner des conclusions sur l'évolution de la robustesse des réseaux aux modes d'itération, lorsque le nombre d'éléments qu'ils comportent augmente. Nous étudions ensuite le nombre moyen d'attracteurs, ainsi que la taille moyenne des cycles limites observés pour les RABS.

### 3.1.1 Classes de comportement dynamique

Nous avons décrit, dans la partie 1.3.2, 4 classes de comportement dynamique pour les réseaux d'automates. Chaque réseau d'automates appartient à l'une de ces classes. Elles sont définies en fonction des attracteurs des réseaux et des modes d'itération avec lesquels les réseaux sont itérés. Nous pouvons ordonner ces 4 comportements dynamiques types, dans le cadre de la robustesse des réseaux face à un changement de mode d'itération. En effet, nous avons déjà expliqué que toute fonction assurée par un réseau de régulation était décrite par un attracteur du réseau d'automates qui le modélise. Si les attracteurs du RABS modèle sont modifiés avec un changement de mode d'itération, alors, cela signifie qu'un changement de synchronisme dans l'expression génétique entraîne une modification de la fonction du réseau. De ce point de vue, les réseaux de la classe de comportement *Fi*, qui admettent uniquement des points fixes quel que soit le mode d'itération avec lequel ils sont itérés, sont donc les réseaux les plus robustes. Nous rappelons en effet que les points fixes d'un réseau d'automates ne dépendent pas du mode d'itération. A l'opposé, les réseaux de la classe *Ev* sont considérés comme étant les plus sensibles aux modes d'itération. Pour ces réseaux, on voit apparaître et disparaître des cycles limites en fonction du mode d'itération choisi.

Il nous semble donc tout à fait pertinent d'étudier la répartition des RABS dans chacune des 4 classes de comportement dynamique. S'il s'avère qu'une grande majorité des RABS se trouve dans la classe *Fi*, alors la grande majorité des RABS est robuste à une perturbation de la façon d'itérer. En revanche, si les classes de comportement les moins robustes, comme *Ev*, se trouvent être les plus représentées, alors l'étude de la robustesse des réseaux à un changement de mode d'itération prend tout son sens.

Afin d'étudier cette répartition des RABS dans les classes de comportement dynamique, nous avons simulé de façon exhaustive l'ensemble de tous les réseaux ayant 3 éléments ou moins. Le nombre de réseaux correspondant à chaque type de comportement est donné dans le tableau 3.1.

Pour les réseaux de taille supérieure ou égale à 4, nous avons simulé la dynamique d'un échantillon de la population totale. Cet échantillon est obtenu par la méthode décrite dans la partie 2.4. Nous avons simulé la dynamique complète de

TAB. 3.1 – Nombre de RABS à $n$ éléments pour chacune des 4 classes de comportement dynamique, pour $1 \leq n \leq 3$.

| $n$ | *Cy* | *Fi* | *Mi* | *Ev* | Total |
|---|---|---|---|---|---|
| 1 | 1 | 3 | 0 | 0 | 4 |
| 2 | 34 | 54 | 8 | 9 | 105 |
| 3 | 71 940 | 59 245 | 22 836 | 34 947 | 188 968 |

$10^8$ RABS de taille 4, $10^7$ RABS de taille 5 et de taille 6, et enfin, $1,5 \times 10^6$ RABS de taille 7. Ces nombres peuvent paraître dérisoires face au nombre total de réseaux pour chacune des tailles. Cependant, nous avons observé que si l'on simule, par exemple, 2 échantillons indépendants de $10^8$ RABS de taille 4 pour estimer la proportion de réseaux correspondant à la classe *Cy*, alors les 2 intervalles de confiance obtenus sont quasiment identiques, la différence étant inférieure à la précision des estimations données dans ce document. On admettra donc que ces échantillons sont représentatifs de la population totale. Remarquons enfin que le nombre de réseaux simulés diminue avec la taille de ceux-ci alors qu'à l'inverse, les populations sont de plus en plus grandes. Nous rappelons que le nombre de configurations et donc de transitions à calculer, tout comme le nombre de modes d'itération croît de façon exponentielle avec la taille des réseaux. Les simulations sont de plus en plus coûteuses en temps d'exécution, à mesure que la taille des réseaux simulés augmente. A titre d'exemple, nous observons que la simulation complète d'un réseau de taille 7 est environ 30 fois plus longue que la simulation d'un réseau de taille 6, dans les mêmes conditions de simulation (même machine).

TAB. 3.2 – Proportion de chacune des 4 classes de comportement dynamique dans la population de RABS à $n$ éléments, pour $1 \leq n \leq 7$.

| $n$ | *Cy* (en %) | *Fi* (en %) | *Mi* (en %) | *Ev* (en %) |
|---|---|---|---|---|
| 1 | $25,00$ | $75,00$ | $0,00$ | $0,00$ |
| 2 | $32,38$ | $51,43$ | $7,62$ | $8,57$ |
| 3 | $38,07$ | $31,35$ | $12,08$ | $18,49$ |
| 4 | $41,52 \pm 0,01$ | $15,79 \pm 0,01$ | $13,80 \pm 0,01$ | $28,88 \pm 0,01$ |
| 5 | $43,94 \pm 0,03$ | $5,94 \pm 0,01$ | $14,17 \pm 0,02$ | $35,95 \pm 0,03$ |
| 6 | $45,75 \pm 0,03$ | $1,52 \pm 0,01$ | $13,66 \pm 0,02$ | $39,07 \pm 0,03$ |
| 7 | $47,14 \pm 0,08$ | $0,25 \pm 0,08$ | $12,66 \pm 0,05$ | $39,95 \pm 0,08$ |

Le tableau 3.2 donne la proportion de chacune des 4 classes de comportement dynamique dans la population de RABS à $n$ éléments, pour $1 \leq n \leq 7$. Cette proportion est exacte pour $n \leq 3$ et estimée pour $4 \leq n \leq 7$. La figure 3.1 reprend les valeurs du tableau 3.2 et les présente graphiquement, afin de mieux visualiser l'évolution de la proportion de chaque classe de comportement, en fonction de la

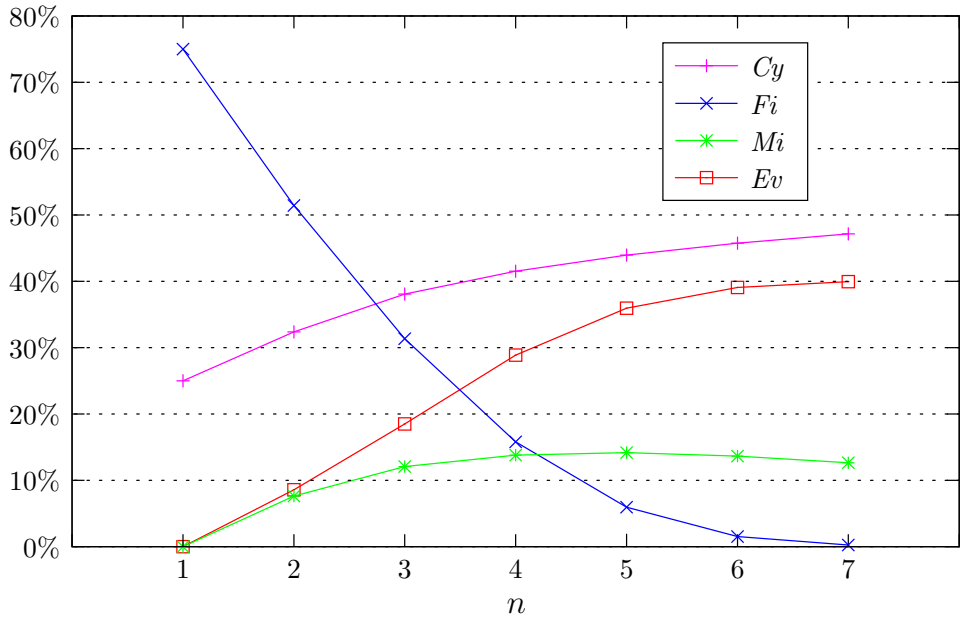



FIG. 3.1 – Evolution de la proportion des 4 classes de comportement dynamique dans la population de RABS à $n$ éléments, en fonction de $n$.

taille des réseaux. Les intervalles de confiance ne sont pas représentés, leur taille les rendant invisibles à l'échelle à laquelle la figure est tracée.

On constate, à partir de ce tableau et de cette figure, que la robustesse des réseaux face à une perturbation de leur mode d'itération diminue avec la taille. En effet, la classe que nous considérons comme étant la plus robuste, *Fi* est de moins en moins représentée. Pour des réseaux de taille 7, on peut même affirmer que la probabilité pour qu'un RABS, tiré au hasard dans la population, n'admette aucun cycle limite, quel que soit le mode d'itération avec lequel il est itéré, est quasi-nulle. A l'inverse, la proportion de la classe *Cy* et, surtout, de la classe *Ev* augmente avec la taille des réseaux. Pour les réseaux de taille 7, ces 2 classes de comportement représentent plus de 85% de la totalité des RABS. La classe *Ev* contient pratiquement 40% des RABS à 7 éléments. Ainsi, pour les réseaux de taille 7, le comportement dynamique est très fortement lié aux modes d'itération, il y est très sensible. Remarquons que la proportion de la classe *Mi* reste de l'ordre de 10 à 15%, lorsque la taille des réseaux passe de 3 à 7 éléments.

Enfin, on voit que les courbes tracées sur la figure 3.1 indiquent une évolution de la proportion de chaque classe dans un unique sens : soit cette proportion augmente avec la taille des réseaux, soit elle décroît (pour la classe *Fi*). Seule la proportion de la classe *Mi* semble lentement décroître à partir de $n = 5$, après avoir augmenté depuis $n = 1$. De plus, on remarque que ces courbes ont tendance à s'infléchir. On peut donc faire l'hypothèse que la proportion de chaque classe de comportement se stabilise à partir d'un certain nombre d'éléments pour les

RABS. D'après ce que l'on observe pour les réseaux de taille inférieure ou égale à 7, on peut supposer que, pour les RABS de taille supérieure, la proportion de réseaux n'admettant aucun cycle limite, quel que soit le mode d'itération, est si faible que l'on peut la considérer comme nulle. Nous reviendrons sur ce point dans le paragraphe suivant. Nous pouvons également émettre l'hypothèse que les classes *Cy* et *Ev* sont toujours très représentées pour les RABS de taille supérieure à 7, et que leur pourcentage semble converger vers une valeur asymptotique (en $n$, taille des RABS), en particulier la classe *Ev*, dont l'évolution semble être sigmoïdale. La classe *Mi* nous pose en revanche un problème. Nous ne pouvons émettre aucune hypothèse concernant l'évolution de la proportion de cette classe : continue-t-elle à décroître jusqu'à atteindre une proportion quasi-nulle comme pour la classe *Fi*, ou bien se stabilise-t-elle autour de 10%, par exemple ?

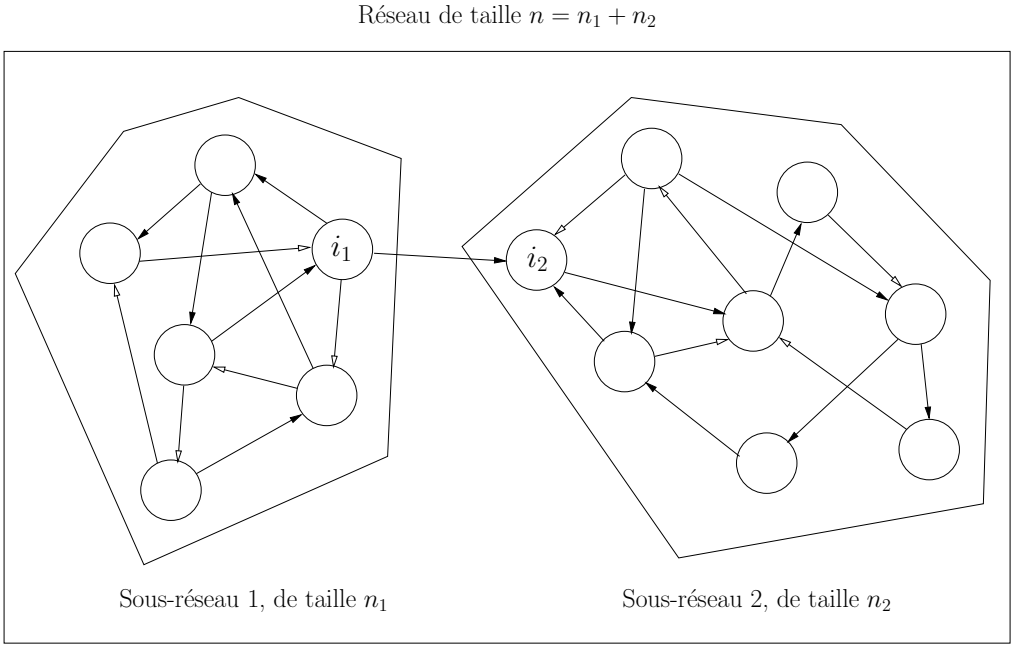



FIG. 3.2 – Décomposition d'un réseau de taille $n$ en 2 sous-réseaux de tailles respectives $n_1$ et $n_2$, chacun constituant une composante fortement connexe, la première composante étant un élément source pour la seconde.

Nous avons affirmé que la proportion de RABS appartenant à la classe de comportement *Fi* pouvait être considérée comme nulle pour des tailles de réseaux supérieures à 7. Cette proportion n'est pas nulle à strictement parler puisque, parmi la population de RABS à $n$ éléments, on peut trouver des RABS dont le graphe d'interaction est par exemple tel que sur la figure 3.2. Ce graphe est divisé en 2 composantes fortement connexes, représentant chacune un RABS de taille inférieure strictement à $n$. Ces 2 sous-réseaux peuvent être des réseaux de la classe *Fi* et être tels que pour chacun des points fixes atteints par le sous-réseau

1, l'état de l'élément $i_1$ du sous-réseau 1, en interface avec le sous-réseau 2, est 0. Dans ce cas, au niveau des attracteurs, l'élément $i_1$ n'a aucun effet sur l'élément $i_2$ du sous-réseau 2. Ainsi, les attracteurs du sous-réseau 1 n'ont aucune influence sur les attracteurs du sous-réseau 2. Dans un tel cas, le réseau total est alors de la classe *Fi*. Le nombre de cas semblables à celui-ci est cependant extrêmement rare, puisque nous avons vu, dans la partie 2.3, que la très grande majorité des RABS de grande taille avaient des graphes d'interaction complets.

Nous observons une baisse de la robustesse moyenne des RABS avec une augmentation de leur taille. Nous nous intéressons aux RABS dans le cadre de la modélisation de réseaux de régulation biologique. Pour ces réseaux, une baisse de la robustesse signifie donc une perte de fonctionnalité. Nous avons vu précédemment que la connectivité moyenne des graphes d'interaction des RABS augmentait avec la taille de ceux-ci. Ainsi, en simulant des réseaux de taille de plus en plus importante, on simule dans le même temps des réseaux de connectivité de plus en plus importante. Or, pour les réseaux de régulation biologique, on observe une connectivité moyenne faible, quelle que soit la taille du réseau. On peut donc s'interroger légitimement sur l'importance relative de la connectivité des RABS vis-à-vis de la perte de robustesse de ces derniers.

Pour réponde à ces interrogations, nous avons travaillé sur des RABS de connectivité relativement faible. Pour cela, on construit les RABS à $n$ éléments, non plus à partir des fonctions booléennes à seuil à $n$ arguments, mais à partir des fonctions booléennes à seuil à $k < n$ arguments. Ainsi, chaque élément du réseau ne peut être régulé que par un nombre maximum $k$ d'éléments du même réseau. La connectivité des RABS ainsi construits est donc au maximum de $k$. Nous avons réalisé 2 séries de simulations. La première série concerne des RABS de taille $n = 6$ fixée, que l'on construit à partir des fonctions booléennes à seuil à $k$ arguments, avec $3 \leq k \leq 5$. La deuxième série de simulations concerne des RABS de taille $n$, avec $4 \leq n \leq 7$, construits à partir des fonctions booléennes à seuil à $k = 3$ arguments. Pour les 2 séries de simulations, on s'assure que les réseaux des échantillons testés sont tous connexes (le graphe non orienté, sous-jacent au graphe d'interaction du réseau est connexe). En effet, dans le cadre de la modélisation de réseaux de régulation génétique, un réseau est toujours connexe, sinon, on considère qu'il s'agit de 2 réseaux. Ce problème de connexité ne se pose pas, lorsque l'on construit les RABS de taille $n$ à partir de fonctions booléennes à seuil à $n$ arguments. Nous avons vu, en effet, que, dans ce cas, la connectivité était de l'ordre de $n$ et que l'on avait donc des graphes d'interaction quasi-complets. En revanche, en diminuant le nombre d'arguments des fonctions booléennes à seuil, on diminue le nombre de voisins qui régulent chacun des éléments du réseau. On peut alors se poser le problème de connexité du réseau.

Afin de réaliser les simulations sur les RABS à $n$ éléments, construits à partir des fonctions booléennes à $k$ arguments, $k < n$, il nous faut connaître la structure de la population de ces RABS. Cette structure nous est nécessaire pour assurer

que l'échantillon de réseaux sur lesquels les simulations seront effectuées est bien représentatif de la population. Dans le cas $n \leq 5$, on peut utiliser la formule de Burnside pour déterminer la structure exacte de la population de RABS à $n$ éléments. Dans le cas $n \geq 6$, nous ne pouvons pas réaliser le calcul de la structure exacte par la formule de Burnside. Ce calcul s'avère être trop long. Nous avons vu dans la partie 2.3 que l'on pouvait définir aisément la structure de 2 populations de RABS à $n$ éléments, appelées populations minorante et majorante. La population minorante nous donne un minorant du nombre de RABS à $n$ éléments, que nous avions noté $Mino_n$. La population majorante fournit un majorant de ce nombre, que nous avions noté $Majo_n$. On note ici $m_n^k$ (resp. $M_n^k$) le nombre de RABS contenus dans la population minorante (resp. majorante), lorsque l'on construit les RABS à $n \geq 6$ éléments à partir de parties génératrices des classes d'équivalence des fonctions booléennes à seuil à $k \leq n$ arguments ($Mino_n = m_n^n$ et $Majo_n = M_n^n$).

TAB. 3.3 – Minorant ($m_6^k$) et majorant ($M_6^k$) du nombre de RABS à 6 éléments construits à partir de fonctions booléennes à seuil à $k$ arguments, pour $3 \leq k \leq 6$.

| $k$ | $m_6^k$ | $M_6^k$ |
|---|---|---|
| 3 | $1,02 \times 10^{16}$ | $3,86 \times 10^{16}$ |
| 4 | $2,71 \times 10^{23}$ | $3,51 \times 10^{23}$ |
| 5 | $3,44 \times 10^{31}$ | $3,53 \times 10^{31}$ |
| 6 | $1,60 \times 10^{40}$ | $1,60 \times 10^{40}$ |

Le tableau 3.3 donne le nombre de RABS à $n = 6$ éléments, construits à partir de fonctions booléennes à $3 \leq k \leq 6$ arguments pour la population minorante ($m_6^k$) et pour la population majorante ($M_6^k$). Alors que le nombre de RABS à 6 éléments construits à partir des fonctions booléennes à seuil à $k = 6$ arguments est quasiment identique entre la population minorante et la population majorante (ce que nous avions remarqué dans la partie 2.3), ce n'est pas le cas pour les RABS construits à partir de fonctions à $k < 6$ arguments. On ne peut donc plus supposer que ces populations sont quasiment identiques. Ainsi, nous ne pouvons plus nous contenter de réaliser les simulations à partir de la population majorante. Pour $3 \leq k \leq 5$, les simulations sont faites sur les réseaux de la population minorante et sur les réseaux de la population majorante. Si les résultats observés sont très peu différents entre ces 2 populations, alors on peut supposer qu'ils sont également valables pour la population exacte de tous les RABS non équivalents (il faudrait que la population de RABS soit vraiment très particulière pour que ce ne soit pas le cas). Pour cette série de simulations, quelle que soit la valeur de $k$, la taille de l'échantillon est $10^5$. Le tableau 3.4 donne la proportion de chaque classe de comportement dynamique dans la population des RABS à 6 éléments, construits à partir de fonctions booléennes à $3 \leq k \leq 5$ arguments, pour la population minorante (ligne m) et pour la population majorante (ligne M). Nous rappelons

également la proportion de chaque classe pour $k = 6$ en dernière ligne du tableau.

TAB. 3.4 – Proportion de chacune des 4 classes de comportement dynamique dans la population des RABS à 6 éléments, construits à partir de fonctions booléennes à $3 \leq k \leq 6$ arguments (la ligne $m$ correspond à la population minorante, la ligne $M$ à la population majorante).

| $k$ | | *Cy* (en %) | *Fi* (en %) | *Mi* (en %) | *Ev* (en %) |
|---|---|---|---|---|---|
| 3 | m | $44,07 \pm 0,31$ | $11,75 \pm 0,20$ | $13,75 \pm 0,21$ | $30,43 \pm 0,29$ |
| | M | $43,58 \pm 0,31$ | $11,41 \pm 0,20$ | $13,92 \pm 0,21$ | $31,09 \pm 0,29$ |
| 4 | m | $43,20 \pm 0,31$ | $9,99 \pm 0,19$ | $14,13 \pm 0,22$ | $32,67 \pm 0,29$ |
| | M | $42,95 \pm 0,31$ | $10,05 \pm 0,19$ | $13,91 \pm 0,21$ | $33,09 \pm 0,29$ |
| 5 | m | $44,22 \pm 0,31$ | $4,58 \pm 0,13$ | $14,54 \pm 0,22$ | $36,67 \pm 0,30$ |
| | M | $44,29 \pm 0,31$ | $4,72 \pm 0,13$ | $14,18 \pm 0,22$ | $36,81 \pm 0,30$ |
| 6 | | $45,75 \pm 0,03$ | $1,52 \pm 0,01$ | $13,66 \pm 0,02$ | $39,07 \pm 0,03$ |

On constate que les résultats pour les populations majorantes et minorantes sont très proches, quel que soit le nombre d'arguments $k$ des fonctions booléennes à seuil utilisées. On considère donc que ces résultats sont valables pour la population exacte de tous les RABS non équivalents. La figure 3.3 présente graphiquement l'évolution de la proportion de chaque classe de comportement dynamique, en fonction du nombre d'arguments $k$ des fonctions booléennes à seuil utilisées pour construire les RABS à 6 éléments. Pour $3 \leq k \leq 5$, la valeur utilisée est celle correspondant à la population majorante.

D'après le tableau 3.4 et la figure 3.3, pour un nombre $n$ d'éléments donné (ici $n = 6$), l'augmentation de la connectivité a pour effet de faire diminuer la robustesse moyenne des réseaux. En effet, on remarque que la proportion de la classe de comportement *Fi* baisse lorsque le nombre d'arguments $k$ des fonctions booléennes à seuil utilisées pour construire les réseaux augmente. A l'opposé, la proportion des classes *Cy*, et surtout *Ev* augmente avec la connectivité. Toujours en considérant que la classe *Fi* est la classe des réseaux les plus robustes, et *Ev*, celle des réseaux les plus sensibles aux changements de mode d'itération, on en déduit que la robustesse moyenne des RABS à 6 éléments, diminue lorsque la connectivité augmente. Ici encore, la proportion de la classe *Mi* semble commencer à décroître pour $k \geq 5$.

En fixant la taille des réseaux et en faisant évoluer leur connectivité, on retrouve les comportements observés précédemment : l'augmentation de la connectivité a tendance à faire diminuer la robustesse des réseaux aux perturbations des modes d'itération.

Dans une deuxième série de simulations, nous étudions le rôle de la connectivité et de la taille des RABS sur leur robustesse, en fixant le nombre d'arguments des fonctions booléennes à seuil. A partir des fonctions booléennes à seuil

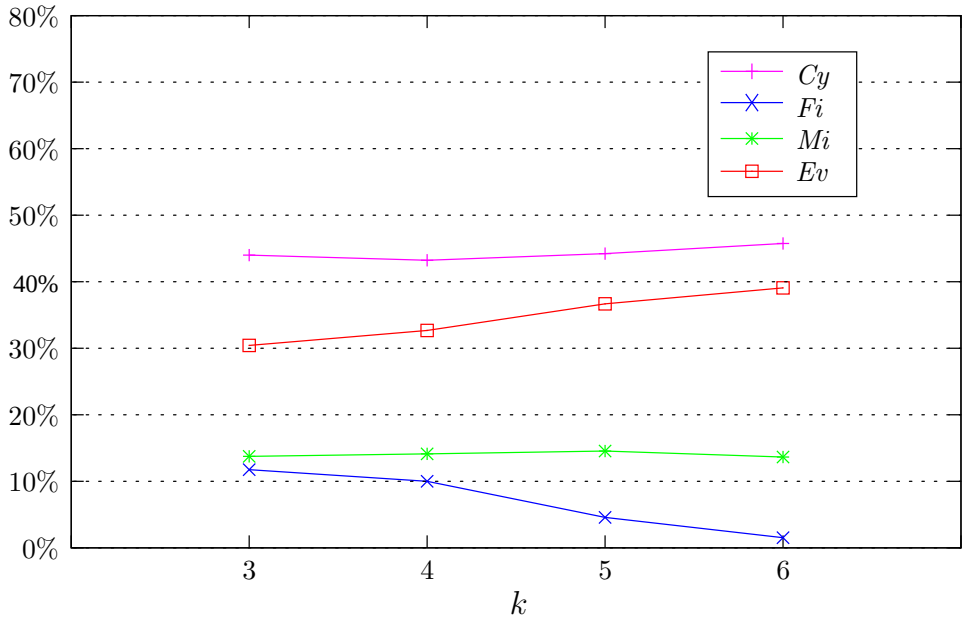



FIG. 3.3 – Evolution de la proportion des 4 classes de comportement dynamique dans la population de RABS à 6 éléments construits à partir de fonctions à seuil à $k$ arguments, en fonction de $k$.

à $k = 3$ arguments, nous construisons des RABS à $n$ éléments, $n$ variant de 4 à 7. Ces fonctions à 3 arguments nous permettent de construire des réseaux dont la connectivité est de l'ordre de 3 : pour les RABS de taille 3, la connectivité moyenne est de $2,60$, comme indiqué dans le tableau 2.6 ; pour les RABS de taille 7, construits à partir des fonctions booléennes à seuil à 3 arguments, la connectivité moyenne est de $3,00$. Ces valeurs sont proches de celles observées pour les réseaux de régulation génétique, qui sont comprises entre $\frac{3}{2}$ et 3 [45, 48]. Comme pour la série de simulations précédentes, nous devons connaître la structure de la population des RABS pour en construire un échantillon représentatif. Ici encore, pour $4 \leq n \leq 5$, on peut déterminer cette structure de population à partir de la formule de Burnside, et pour $6 \leq n \leq 7$, nous utilisons les populations minorante et majorante. Le tableau 3.5 présente le nombre de RABS à $n$ éléments, construits à partir des fonctions booléennes à seuil à 3 arguments : pour $3 \leq n \leq 5$, le nombre exact, noté $N_n^3$, est donné, et pour $6 \leq n \leq 7$, un minorant, $m_n^3$, et un majorant, $M_n^3$, de ce nombre sont donnés.

On remarque que, pour $6 \leq n \leq 7$, le nombre de RABS à $n$ éléments, construits à partir des fonctions booléennes à 3 arguments, est très différent entre les populations minorante et majorante. Comme pour la précédente série de simulations, nous devons donc réaliser un échantillon à partir de la population minorante, et un échantillon à partir de la population majorante, pour les RABS de taille supérieure ou égale à 6. Si les résultats obtenus à partir de ces 2

TAB. 3.5 – $N_n^3$, nombre de RABS à $n$ éléments construits à partir de fonctions booléennes à seuil à $k = 3$ arguments, pour $3 \leq n \leq 5$ ; pour $6 \leq n \leq 7$, on donne un minorant ($m_n^3$) et un majorant ($M_n^3$) de ce nombre.

| $n$ | $m_n^3$ | $N_n^3$ | $M_n^3$ |
|---|---|---|---|
| 3 | - | 188 968 | - |
| 4 | - | $5,97 \times 10^8$ | - |
| 5 | - | $2,94 \times 10^{12}$ | - |
| 6 | $1,02 \times 10^{16}$ | - | $3,86 \times 10^{16}$ |
| 7 | $7,28 \times 10^{19}$ | - | $4,71 \times 10^{20}$ |

échantillons sont proches, alors on considère qu'ils sont valables également pour la population exacte. Les simulations sont effectuées sur des échantillons de $10^5$ RABS. Le tableau 3.6 donne la proportion de chaque classe de comportement dynamique dans la population de RABS à $n$ éléments, construits à partir des fonction booléennes à seuil à 3 arguments, pour $3 \leq n \leq 5$. Pour les cas $n = 6$ et $n = 7$, les proportions sont données pour la population minorante (ligne m) et pour la population majorante (ligne M). La première ligne du tableau, qui correspond à $n = 3$, donne les proportions exactes.

TAB. 3.6 – Proportion de chacune des 4 classes de comportement dynamique dans la population des RABS à $3 \leq n \leq 7$ éléments, construits à partir de fonctions booléennes à $k = 3$ arguments (la ligne $m$ correspond à la population minorante, la ligne $M$ à la population majorante).

| $n$ | | *Cy* (en %) | *Fi* (en %) | *Mi* (en %) | *Ev* (en %) |
|---|---|---|---|---|---|
| 3 | | $38,07$ | $31,35$ | $12,08$ | $18,49$ |
| 4 | | $40,66 \pm 0,10$ | $21,64 \pm 0,08$ | $12,94 \pm 0,07$ | $24,76 \pm 0,08$ |
| 5 | | $42,36 \pm 0,10$ | $15,60 \pm 0,07$ | $13,48 \pm 0,07$ | $28,56 \pm 0,09$ |
| 6 | m | $44,07 \pm 0,31$ | $11,75 \pm 0,20$ | $13,75 \pm 0,21$ | $30,43 \pm 0,29$ |
| | M | $43,58 \pm 0,31$ | $11,41 \pm 0,20$ | $13,92 \pm 0,21$ | $31,09 \pm 0,29$ |
| 7 | m | $42,47 \pm 0,31$ | $17,66 \pm 0,24$ | $12,09 \pm 0,20$ | $27,78 \pm 0,28$ |
| | M | $41,95 \pm 0,31$ | $17,70 \pm 0,24$ | $12,32 \pm 0,21$ | $28,03 \pm 0,28$ |

Comme pour la série de simulations précédente, on constate que, pour $6 \leq n \leq 7$, les résultats obtenus pour les populations minorante et majorante sont très proches. On peut donc considérer que ces résultats sont également valables pour la population exacte de tous les RABS non équivalents. La figure 3.4 présente graphiquement l'évolution de la proportion de chaque classe de comportement dynamique, en fonction de la taille $n$ des RABS construits à partir des fonctions booléennes à seuil à 3 arguments. Pour $6 \leq n \leq 7$, la valeur utilisée pour le graphique correspond à celle de la population majorante.

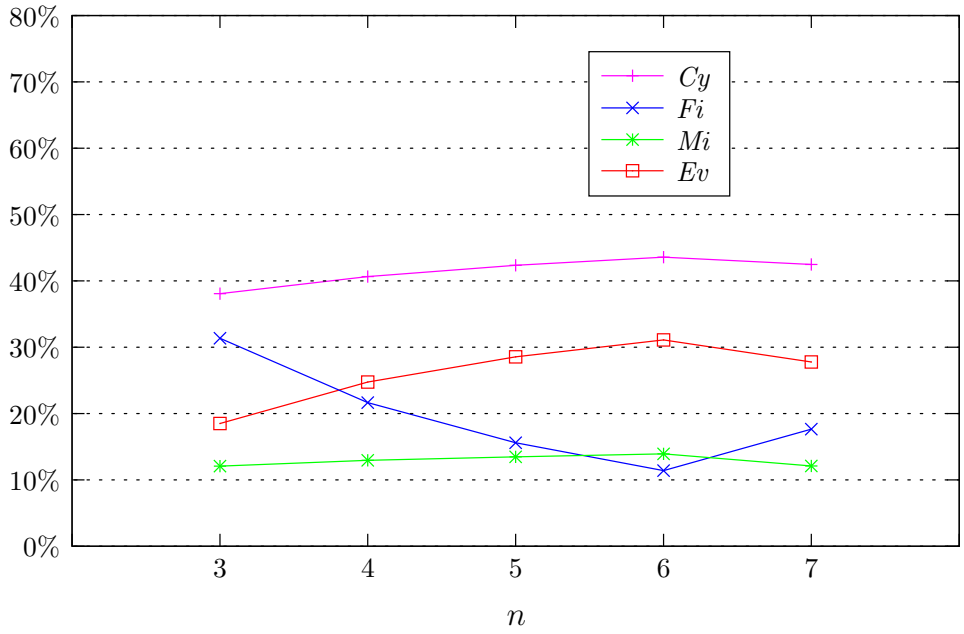



FIG. 3.4 – Evolution de la proportion des 4 classes de comportement dynamique dans la population de RABS à $n$ éléments construits à partir de fonctions booléennes à seuil à $k = 3$ arguments, en fonction de $n$.

D'après le tableau 3.6 et la figure 3.4, on constate que, pour une connectivité donnée (les réseaux simulés sont à peu près tous de même connectivité, quelle que soit leur taille), l'augmentation de la taille des RABS a tendance à faire diminuer la robustesse de ceux-ci. On retrouve, comme dans les simulations précédentes, une décroissance de la proportion de la classe *Fi*, associée à une croissance de la proportion des classes *Cy* et *Ev*. Cependant, ce phénomène n'est observé que pour les réseaux de taille inférieure ou égale à 6. Entre les RABS de taille 6 et 7, on observe, au contraire, que la proportion de la classe *Fi* croît, alors que celle des autres classes de comportement décroît (logiquement, la somme des proportions étant toujours égale à 1). La proportion de la classe *Fi* pour les RABS de taille 7 est même supérieure à celle pour les RABS de taille 5. On remarque, par ailleurs, que les proportions des classes de comportement dynamique *Fi* et *Ev* semblent être liées : pour les tailles inférieures à 6, la proportion de la classe *Ev* semble d'autant plus croître avec le nombre d'éléments des RABS que la proportion de la classe *Fi* décroît ; entre $n = 6$ et $n = 7$, la proportion de *Fi* croît fortement, et la classe *Ev* est la classe dont la proportion connaît la plus forte baisse. Nous pouvions déjà observer ce phénomène à propos des simulations précédentes. Les classes *Fi* et *Ev* sont celles dont les proportions évoluent le plus fortement avec des changement de taille de réseaux ou de connectivité. Lorsque la proportion de l'une augmente, celle de l'autre diminue.

Nous avons vu, dans le paragraphe précédent, qu'entre des RABS à 6 élé-

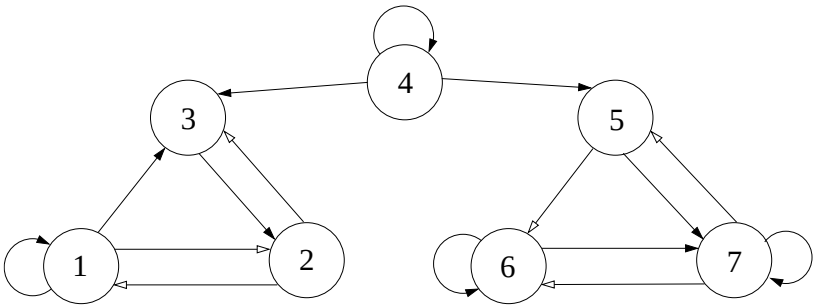



FIG. 3.5 – Graphe d'interaction d'un RABS à 7 éléments, construit à partir de fonctions booléennes à seuil à 3 arguments, et appartenant à la classe *Fi*.

ments, de connectivité égale à 3 et des RABS à 7 éléments, de même connectivité, la proportion de la classe *Fi* augmentait. En cela, nous pouvons dire que la robustesse moyenne (toujours aux changement de modes d'itération) des RABS à 7 éléments est plus importante que celle des RABS à 6 éléments, lorsque la connectivité est faible. Nous allons tenter de donner une explication de ce phénomène qui nous paraît plausible, à l'aide d'un exemple. Le RABS dont le graphe d'interaction est présenté figure 3.5 est un RABS à 7 éléments, construit à partir des fonctions booléennes à 3 arguments. Chacun de ses éléments, représenté par un noeud du graphe est en effet régulé par, au plus, 3 autres éléments. La matrice d'interaction $\mathbf{W}$ et le vecteur seuil $\boldsymbol{\theta}$ qui définissent ce réseau sont donnés ci-après :

$$\mathbf{W} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & -1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix} \qquad \boldsymbol{\theta} = \begin{pmatrix} -1 \\ 0 \\ 0 \\ 0 \\ -1 \\ 0 \\ 1 \end{pmatrix}$$

C'est un réseau de la classe *Fi*, puisque, quel que soit le mode d'itération utilisé, il admet uniquement 2 points fixes qui sont 1010100 et 1011101. On peut remarquer, à partir du graphe ou de la matrice d'interaction, que ce RABS est constitué de 2 sous-réseaux, que nous appellerons modules. Le premier module regroupe les éléments 1, 2 et 3, alors que le second module regroupe les éléments 5, 6 et 7. Pris séparément, ces 2 modules sont 2 RABS de taille 3. Les 2 appartiennent à la classe *Fi*. Ainsi, le RABS à 7 éléments est constitué de 2 modules qui peuvent évoluer de façon indépendante, en fonction de la valeur de l'élément qui les relie. Une telle configuration ne peut être trouvée parmi les RABS de taille inférieure à 7. En effet, si le nombre d'éléments était ici de 6, c'est-à-dire si l'élément 4 n'existait pas dans le réseau, alors, pour avoir un réseau connexe, les 2 modules seraient liés au moyen d'une interaction entre un noeud d'un module

et un noeud de l'autre. Les 2 modules ne pourraient donc plus évoluer indépendamment, et selon le type d'interaction, on pourrait observer des cycles limites. Dans les réseaux de taille 7, on peut associer de diverses manières des modules de taille 3, qui évoluent de façon indépendante selon leur propre dynamique. Parmi les réseaux de taille 3, un peu moins du tiers sont des réseaux de la classe *Fi*. En construisant les réseaux de taille 7 à partir de ces modules, on obtient sous certaines conditions des réseaux de la classe *Fi* également.

Nous pensons que c'est cette possibilité de construire des RABS de taille 7 à partir de modules indépendants de taille 3 (voire moins), induite par le fait de conserver une connectivité faible, qui permet d'expliquer que la proportion de la classe *Fi* est plus élevée pour les RABS de taille 7 que pour les RABS de taille 6. Cet exemple illustre le fait que les réseaux de régulation génétique sont constitués de modules ou motifs qui sont reliés entre eux par des noeuds intermédiaires. Cette configuration en modules permettrait aux réseaux de taille suffisamment grande de présenter une plus forte robustesse aux changements de mode d'itération.

L'ensemble des simulations présentées montre que l'accroissement de la taille des RABS s'accompagne d'une baisse de la robustesse moyenne de ces réseaux aux modes d'itération. Cette diminution de la robustesse s'explique notamment par une connectivité des réseaux qui augmente linéairement avec leur taille. Si l'on fixe la connectivité à une valeur faible, à partir d'une certaine taille, les réseaux semblent présenter une plus forte robustesse. Nous avons vu que les réseaux pour lesquels on observe des cycles limites, étaient très représentés, et que leur proportion augmente avec la taille des réseaux. Nous allons présenter quelques statistiques concernant ces cycles limites, et plus généralement, les attracteurs des réseaux.

### 3.1.2 Attracteurs

Une première série de simulations concerne le nombre moyen de points fixes, de cycles limites et d'attracteurs par réseau et par mode d'itération. Le nombre d'attracteurs est égal à la somme du nombre de points fixes et du nombre de cycles limites. Nous calculons ce nombre moyen pour des RABS de taille $3 \leq n \leq 6$, construits à partir de fonction booléennes à seuil à $k = n$ arguments. Pour chaque taille de RABS supérieure à 3, l'échantillon étudié comporte $10^5$ réseaux. Pour les réseaux de taille 3, les nombres d'attracteurs sont calculés sur la population totale. Le tableau 3.7 présente, pour chaque taille de réseau, le nombre moyens de points fixes (ligne "Pts Fixes"), de cycles limites (ligne "Cycle") et d'attracteurs (ligne "Attract") par réseau et par mode d'itération, en fonction de la classe de comportement dynamique, et pour l'ensemble des réseaux (colonne "Total").

On constate tout d'abord que le nombre moyen d'attracteurs des réseaux de la classe *Mi* est toujours supérieur à 2. Ceci vient de la définition même de la classe *Mi* qui regroupe tous les réseaux pour lesquels on observe un cycle limite et un point fixe pour tous les modes d'itération. On constate également que le

TAB. 3.7 – Nombre moyen de points fixes (Pts Fixes), de cycles limites (Cycles) et d'attracteurs (Attract.) par réseau et par mode d'itération, en fonction de la classe de comportement dynamique, pour les RABS à $n$ éléments, avec $3 \leq n \leq 6$.

| $n$ | | *Cy* | *Fi* | *Mi* | *Ev* | Total |
|---|---|---|---|---|---|---|
| 3 | Pts Fixes | 0 | 1,93 | 1,19 | 1,36 | 1,00 |
| | Cycles | 1,15 | 0 | 1,07 | 0,33 | 0,63 |
| | Attract. | 1,15 | 1,93 | 2,26 | 1,69 | 1,63 |
| 4 | Pts Fixes | 0 | 2,30 | 1,35 | 1,56 | 1,00 |
| | Cycles | 1,25 | 0 | 1,17 | 0,37 | 0,80 |
| | Attract. | 1,25 | 2,20 | 2,52 | 1,93 | 1,79 |
| 5 | Pts Fixes | 0 | 2,80 | 1,49 | 1,73 | 1,00 |
| | Cycles | 1,36 | 0 | 1,30 | 0,45 | 0,94 |
| | Attract. | 1,36 | 2,80 | 2,79 | 2,18 | 1,94 |
| 6 | Pts Fixes | 0 | 3,49 | 1,63 | 1,86 | 1,00 |
| | Cycles | 1,47 | 0 | 1,45 | 0,57 | 1,09 |
| | Attract. | 1,47 | 3,49 | 3,08 | 2,43 | 2,09 |

nombre moyen de cycles limites des réseaux de la classe *Ev* est inférieur à 1. Ceci s'explique par le fait que ce nombre est le nombre moyen de cycles limites par réseau et par mode d'itération. Or les réseaux de la classe *Ev* sont caractérisés par le fait de n'avoir des cycles limites que pour certains modes d'itération.

D'un point de vue plus général, on constate que le nombre moyen de points fixes pour les réseaux ayant des points fixes est bien plus élevé que le nombre moyen de cycles limites pour les réseaux ayant des cycles limites. Le nombre moyen de points fixes pour les réseaux de la classe *Fi* croît même plus rapidement que le nombre moyen de cycles limites pour les réseaux de la classe *Cy*, avec la taille des réseaux. Mais si l'on regarde sur l'ensemble de la population, alors le nombre moyen de points fixes par réseau et par mode d'itération ne croît pas avec la taille des réseaux, alors que le nombre moyen de cycles limites croît, jusqu'à dépasser le nombre moyen de points fixes pour les RABS de taille 6. Ceci est dû au fait que la proportion de réseaux ayant des cycles limites (les réseaux des classes *Cy*, *Mi* et *Ev*) augmente avec la taille des réseaux, alors que celle des réseaux ayant des points fixes (les réseaux des classes *Fi*, *Mi* et *Ev*) diminue : on passe de $68,5\%$ des réseaux de taille 3 avec au moins un cycle limite pour un mode d'itération, à $98,5\%$ des réseaux de taille 6 ; pour les points fixes, on passe de $61,9\%$ à $52,9\%$. Le phénomène est encore accentué par le fait que la classe pour laquelle il y a le plus de points fixes en moyenne par réseau, est la classe *Fi*, dont la proportion tend vers 0 lorsque la taille des réseaux augmente. Ainsi, en tirant au sort un réseau dans la population des RABS de taille 6, et en tirant ensuite au sort un mode d'itération, le nombre moyen de cycles limites observés lorsque l'on itère le réseau sera plus élevé que le nombre moyen de points fixes.

Il faut tout de même noter qu'il est étonnant et intéressant de constater que le nombre moyen de points fixes par réseau, pour l'ensemble de la population des RABS, est constant en fonction de la taille des réseaux. Nous notons, d'après nos simulations, une très faible croissance de ce nombre moyen de points fixes par réseau, avec l'augmentation de la taille des réseaux : l'augmentation est de l'ordre de $4 \times 10^{-3}$ entre les tailles 3 et 6. Cette croissance est trop faible pour que nous puissions être absolument certains qu'elle n'est pas uniquement due à l'échantillonnage.

Enfin, concernant le nombre moyen d'attracteurs par réseau et par mode d'itération dans la population globale, les résultats obtenus sont un peu plus faibles que ceux obtenus pour les réseaux booléens aléatoires de Kauffman de petite taille, par Gershenson dans [88]. Par exemple, pour des réseaux booléens aléatoires de taille 5, construits à partir de fonctions booléennes à 5 arguments, Gershenson trouve un nombre moyen d'attracteurs par réseau et par mode d'itération compris en 2,25 et 2,5, selon la classe de réseaux étudiée. Ici, nous trouvons un nombre moyen d'attracteurs par réseau et par mode d'itération de 1,94. Nous supposons que l'on peut expliquer cette différence par le type de fonctions de transition locale utilisées pour construire les réseaux : booléennes dans le cas des réseaux de Kauffman, et booléennes à seuil dans notre cas. Enfin, on observe que le nombre moyen d'attracteurs pour des RABS de taille $n$ est de l'ordre de $\sqrt{n}$, pour $3 \leq n \leq 6$. Ceci est en contradiction avec les résultats récents concernant les réseaux booléens aléatoires [54, 55, 56, 57], qui montrent que le nombre de cycles limites croît plus fortement que la taille des réseaux. Ces résultats sont cependant valables pour des réseaux de grande taille ($> 10$). Dans notre cas, la taille des RABS est trop faible pour conclure sur le fait que la croissance du nombre moyen d'attracteurs est de type linéaire ou si ce nombre est en $O(\sqrt{n})$ (cf. également [53] pour un résultat concernant cette approximation asymptotique dans le cas des RABS).

Les RABS à $n$ éléments semblent avoir moins d'attracteurs que les réseaux d'automates booléens généraux à $n$ éléments. Le nombre moyen de cycles limites par réseau et par mode d'itération est plus faible pour les petites tailles que le nombre moyen de points fixes. A partir de $n = 6$, les cycles limites sont plus nombreux du fait que la proportion de réseaux avec cycles limites est bien plus importante que la proportion de réseaux avec points fixes.

Dans une deuxième série de simulations, nous nous sommes intéressés à la taille moyenne des cycles limites observés. Pour ces simulations, nous n'avons pas pris en compte les réseaux de la classe *Fi*, puisqu'ils n'ont pas de cycle limite. Les échantillons sont encore de taille $10^5$ pour les réseaux de taille 4 à 6. Pour les réseaux de taille 3, les longueurs moyennes des cycles limites correspondent aux longueurs observées sur l'ensemble de la population. Le tableau 3.8 présente la longueur moyenne des cycles limites, pour les RABS de taille 3 à 6, en fonction de la classe de comportement dynamique, et pour l'ensemble des réseaux avec cycles limites (colonne "Total")

On constate tout d'abord que, quelle que soit la taille des réseaux, les réseaux

TAB. 3.8 – Longueur moyenne des cycles limites pour les RABS de taille $3 \leq n \leq 6$, en fonction de la classe de comportement dynamique.

| $n$ | *Cy* | *Mi* | *Ev* | Total |
|---|---|---|---|---|
| 3 | 2,23 | 2,06 | 2,11 | 2,17 |
| 4 | 2,32 | 2,10 | 2,22 | 2,25 |
| 5 | 2,41 | 2,13 | 2,34 | 2,34 |
| 6 | 2,51 | 2,16 | 2,46 | 2,44 |

de la classe *Cy* ont en moyenne des cycles plus longs que ceux de la classe *Ev*, et que ceux de la classe *Mi*. D'après le tableau 3.7, ces cycles plus longs sont également en moyenne plus nombreux pour les réseaux de la classe *Cy* que pour ceux de la classe *Mi*. Enfin, la longueur moyenne des cycles limites des réseaux de la classe *Mi* augmente faiblement avec la taille des réseaux.

D'un point de vue général, on remarque que la longueur moyenne des cycles limites reste assez faible, même pour des RABS de taille 6. Les cycles limites de longueur 2 sont majoritaires pour toutes les tailles de réseau entre 3 et 6. Ce résultat concorde avec les résultats obtenus par Samuelsson et Troein [55], pour les réseaux booléens aléatoires. En revanche, la longueur moyenne des cycles limites observées est plus faible pour les RABS que pour les réseaux booléens aléatoires : nous observons une longueur moyenne de 2,44 pour les RABS à 6 éléments, alors que l'on peut estimer que la longueur moyenne des cycles observés pour les réseaux booléens aléatoires à 6 éléments est de l'ordre de 2,7-2,8, toujours d'après [55].

Nos simulations montrent que les RABS ont un nombre moyen d'attracteurs qui augmente assez faiblement avec la taille des réseaux. La longueur moyenne des cycles limites des RABS évolue également assez lentement avec la taille des réseaux. Ces 2 points différencient les RABS des réseaux booléens aléatoires, pour lesquels le nombre moyen d'attracteurs, tout comme la longueur moyenne des cycles limites, sont plus importants (pour une même taille de réseau) et augmentent plus fortement.

## 3.2 Etude de la classe de comportement *Ev*

Dans la partie précédente, nous avons montré que, lorsque la taille des RABS augmentait, ceux-ci devenaient plus sensibles à un changement de mode d'itération. Nous avons notamment observé que la classe de comportement *Ev* était de plus en plus représentée parmi les RABS. Cette classe est très particulière, puisqu'elle regroupe les réseaux dont le comportement dynamique est le plus variable en fonction des modes d'itération. On observe l'apparition et la disparition de cycles limites en modifiant le synchronisme des éléments du réseaux. Afin de

mieux comprendre ce phénomène, nous définissons une relation d'ordre pour l'ensemble des modes d'itération blocs-séquentiels et, au moyen de simulations, nous étudions le lien entre cette relation d'ordre et les cycles limites des RABS de la classe *Ev*.

Dans la partie 1.3.1, nous avons montré que l'on pouvait associer tout mode d'itération blocs-séquentiel pour un réseau d'automates à $n$ éléments, à une unique partition ordonnée de l'ensemble $\{1, \cdots, n\}$. Nous allons définir la relation d'ordre sur les modes d'itération blocs-séquentiels à partir d'une relation d'ordre sur les partitions ordonnées.

**Définition 3.1** (Relation d'ordre sur les partitions ordonnées)
*Soient $P_1$ et $P_2$ deux partitions ordonnées de l'ensemble $\{1, \cdots, n\}$.*
*A chacune des partitions $P_1$ et $P_2$ est associé un préordre total sur $\{1, \cdots, n\}$.*
*On note $\preceq_1$ (resp. $\preceq_2$) le préordre associé à $P_1$ (resp. $P_2$).*
*On dit que la partition $P_2$ est une partition ordonnée* plus fine *que $P_1$ si, et seulement si :*

$$\forall i, j \in \{1, \cdots, n\}, \;\; i \preceq_1 j \text{ et } j \not\preceq_1 i \;\; \Rightarrow \;\; i \preceq_2 j \text{ et } j \not\preceq_2 i$$

La relation "plus fine" définit bien une relation d'ordre partiel sur l'ensemble des partitions ordonnées de $\{1, \cdots, n\}$. Par exemple, la partition ordonnée $(1)(3)(2,4)$ est plus fine que la partition ordonnée $(1)(2,3,4)$. En revanche, les partitions ordonnées $(1)(3)(2,4)$ et $(1,2)(3,4)$ ne sont pas comparables par la relation "plus fine".

De cette relation d'ordre sur les partitions ordonnées découle naturellement une relation d'ordre sur les modes d'itération blocs-séquentiels.

**Définition 3.2** (Relation d'ordre sur les modes d'itération blocs-séquentiels)
*Soient $M_1$ et $M_2$ deux modes d'itération blocs-séquentiels associés aux partitions ordonnées de $\{1, \cdots, n\}$ $P_1$ et $P_2$.*
*On dit que $M_1$ est un mode d'itération blocs-séquentiel* plus fin *que le mode d'itération $M_2$ si, et seulement si la partition ordonnée $P_1$ est plus fine que $P_2$.*

La relation "plus fine" étant une relation d'ordre partiel sur l'ensemble des partitions ordonnées de $\{1, \cdots, n\}$, la relation "plus fin" est bien une relation d'ordre partiel sur l'ensemble des modes d'itération d'un réseaux d'automates à $n$ éléments. Tout mode d'itération blocs-séquentiel est plus fin que le mode parallèle. Inversement, il n'existe aucun mode d'itération blocs-séquentiel plus fin qu'un mode séquentiel (excepté lui-même).

Notons que Goles et Martinez [72] utilisent déjà cette relation d'ordre sur les modes d'itération blocs-séquentiels, notamment pour l'étude des RABS à matrice d'interaction symétrique.

Munis de leur relation d'ordre partiel respective, l'ensemble des partitions ordonnées de $\{1, \cdots, n\}$ et l'ensemble de modes d'itération blocs-séquentiels pour les réseaux d'automates à $n$ éléments, sont ce que l'on appelle des treillis. On peut représenter chacun d'eux par un diagramme de Hasse. Dans un tel diagramme, tous les éléments de l'ensemble sont ordonnés verticalement en fonction de la

relation d'ordre (les éléments les plus "petits" selon la relation d'ordre sont en bas, les plus "grands", en haut), et les éléments comparables sont reliés entre eux. Afin d'avoir un diagramme lisible, on ne conserve que les liens nécessaires. Par exemple, si $P_1$ est plus fine que $P_2$ et s'il existe $P_3$ plus fine que $P_2$ telle que $P_1$ est plus fine que $P_3$, on ne représente pas le lien entre $P_2$ et $P_1$ ($P_1, P_2$ et $P_3$ représentent des partitions ordonnées de $\{1, \cdots, n\}$). La figure 3.6 présente le diagramme de Hasse correspondant au treillis des modes d'itération blocs-séquentiels pour les automates à 3 éléments. Les modes d'itération parallèle et séquentiels sont indiqués sur le diagramme.

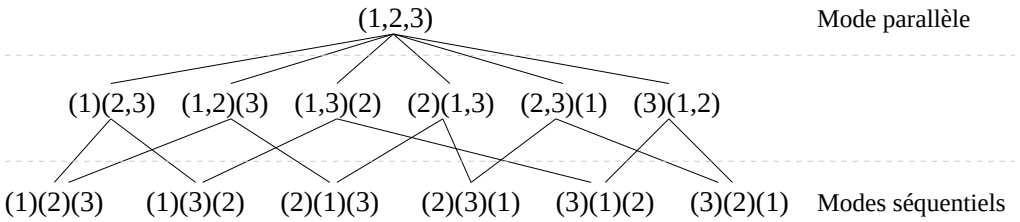



FIG. 3.6 – Représentation du treillis des modes d'itération blocs-séquentiels pour les réseaux d'automates à 3 éléments.

Nous voyons sur la figure 3.6 que l'on obtient un mode d'itération blocs-séquentiel plus fin, à partir d'un mode d'itération donné, en séparant les éléments d'un même bloc du premier mode d'itération en 2 blocs et en conservant l'ordre des blocs. Par exemple, on peut obtenir le mode d'itération $(1)(3)(2)$ à partir du mode $(1)(2,3)$, en séparant les éléments 2 et 3, ou, encore, à partir du mode $(1,3)(2)$, en séparant les éléments 1 et 3. La relation d'ordre sur les modes d'itération est donc liée à la (dé)synchronisation des éléments des réseaux d'automates. A partir de cette relation d'ordre et du diagramme de Hasse qui la représente, nous pouvons regrouper les réseaux d'automates de la classe de comportement dynamique *Ev* en 3 sous-classes de comportement :

- la première sous-classe de comportement englobe tous les réseaux d'automates pour lesquels, s'il n'y a aucun cycle limite observé pour un mode d'itération donné, alors, on n'observe aucun cycle limite pour tout mode d'itération plus fin ; pour de tels réseaux, la disparition des cycles limites se fait en parcourant le diagramme de Hasse de haut en bas ; on appelle donc cette sous-classe, la sous-classe *Down* ;
- la deuxième sous-classe de comportement est celle qui réunit tous les réseaux pour lesquels, s'il n'y a aucun cycle limite observé pour un mode d'itération donné, alors, on n'observe aucun cycle limite pour tout mode d'itération moins fin ; pour de tels réseaux, la disparition des cycles limites se fait en parcourant le diagramme de Hasse de bas en haut ; on appelle donc cette sous-classe, la sous-classe *Up* ;
- enfin, la troisième sous-classe de comportement regroupe tous les réseaux de la classe *Ev*, qui n'appartiennent ni à la sous-classe *Down*, ni à la sous-classe

*Up* ; on appelle donc cette sous-classe, la sous-classe *None*.

D'après cette définition des sous-classes de *Ev*, les réseaux de *Ev* qui n'ont pas de cycle limite dès que l'on désynchronise certains de leurs éléments (toujours les mêmes), sont des réseaux de la sous-classe *Down*. Les réseaux, pour lesquels, au contraire, on note une disparition des cycles limites en synchronisant certains de leurs éléments (toujours les mêmes), sont des réseaux de la sous-classe *Up*. Traditionnellement, on considère que, pour un réseau d'automates donné, les cycles limites disparaissent lorsque l'on désynchronise certains éléments de ce réseau. Ainsi, nous nous attendons à ce que la sous-classe *Up* soit très peu, voire pas représentée. Pour la suite, on désignera les sous-classes *Down*, *Up* et *None* en employant le terme de "classes", afin de simplifier l'écriture et la lecture de ce document.

D'après la définition des classes *Down* et *Up*, on peut faire les 2 remarques suivantes :

- tout réseau de la classe *Down* a un cycle limite pour le mode d'itération parallèle (sinon, il n'a de cycle limite pour aucun des modes d'itération blocs-séquentiels, et n'est donc pas un réseau de la classe *Ev*) ;
- au contraire, tout réseau de la classe *Up* n'a pas de cycle limite pour le mode d'itération parallèle (sinon, il a un ou plusieurs cycle(s) limite(s) pour tous les modes d'itération blocs-séquentiels, et n'est donc pas un réseau de la classe *Ev*).

Nous pouvons également remarquer que, dans le cadre de la modélisation des réseaux de régulation génétique, les classes *Down* et *Up* ne sont pas du tout du même intérêt. Nous avons vu que nous pouvions regrouper deux gènes dans un même bloc d'itération, lorsque nous supposions que les processus d'expression de ces gènes étaient synchrones. Ceci est une simplification de la réalité biologique. En effet, on peut considérer que les gènes sont tous désynchronisés, mais que les intervalles de temps entre les expressions de certains d'entre eux sont tellement courts que l'on peut supposer qu'ils sont synchrones. Pour les réseaux de la classe *Down*, cette simplification n'a pas d'effet important, puisque, en désynchronisant les mises à jour des différents éléments de ces réseaux, on ne voit pas de cycle limite apparaître. En revanche, si l'on modélise un réseau de régulation par un RABS de la classe *Up*, alors la synchronisation des processus d'expression des différents gènes de ce réseau est un élément très important à connaître. En effet, si l'on suppose que les éléments sont synchronisés pour leur mise à jour, alors qu'ils ne le sont pas tout à fait dans la réalité, on peut oublier de prendre en compte un cycle limite, puisque les cycles limites ont tendance à disparaître lorsque l'on synchronise les mises à jour des éléments des RABS de la classe *Up*. A ce titre, les réseaux de cette classe de comportement ne sont pas de bons modèles pour les réseaux de régulation génétique.

Afin de comprendre comment les cycles limites des RABS de la classe de comportement *Ev* apparaissent et disparaissent, nous étudions, à l'aide de simulations, la répartition de ces réseaux dans les différentes classes *Down*, *Up* et *None*, pour des tailles de réseaux comprises entre 2 et 7. Nous ne traitons pas les

réseaux de taille 1, car aucun n'est de la classe *Ev*. Pour les réseaux de tailles 2 et 3, les simulations sont faites sur l'ensemble de la population des RABS. Les 9 RABS de taille 2 qui sont de la classe *Ev* sont tous des réseaux de la classe *Down* (ils ont un unique cycle limite de longueur 2 pour le mode parallèle uniquement). Parmi les 34 947 RABS de taille 3 de la classe *Ev*, 21 729 sont de la classe *Down*, 108 sont de la classe *Up* et les 13 110 restant sont de la classe *None*. Pour les tailles supérieures, nous travaillons sur des échantillons représentatifs de la population des RABS comme pour la première série de simulations de la partie 3.1.1. Pour les RABS à 4 éléments, la taille de l'échantillon est $10^8$, pour les RABS à 5 éléments, la taille est $10^7$, pour les RABS à 6 éléments, la taille est $10^6$ et pour les RABS à 7 éléments, la taille de l'échantillon est $10^5$. Notons enfin que nous donnons la répartition des RABS à $n$ éléments dans les 3 classes de comportement *Down*, *Up* et *None*, sous forme de proportions de la classe *Ev*. Ainsi, les marges de confiance pour les proportions sont donc calculées, non pas à partir de la taille de l'échantillon total, mais à partir du nombre de RABS de la classe *Ev*. Le tableau 3.9 donne la répartition des RABS à $n$ éléments dans les 3 classes de comportement *Down*, *Up* et *None*, pour $2 \leq n \leq 7$.

TAB. 3.9 – Distribution des 3 classes de comportement *Down*, *Up* et *None* pour les RABS de taille $n$, avec $2 \leq n \leq 7$. Les résultats sont donnés sous forme de pourcentage de la classe *Ev*.

| $n$ | *Down* (en %) | *Up* (en %) | *None* (en %) |
|---|---|---|---|
| 2 | 100 | 0 | 0 |
| 3 | $62,18$ | $0,31$ | $37,51$ |
| 4 | $28,87 \pm 0,02$ | $0,21 \pm 0,00$ | $70,93 \pm 0,02$ |
| 5 | $9,72 \pm 0,03$ | $0,08 \pm 0,00$ | $90,20 \pm 0,03$ |
| 6 | $2,41 \pm 0,03$ | $0,02 \pm 0,00$ | $97,57 \pm 0,03$ |
| 7 | $0,25 \pm 0,07$ | $0,00 \pm 0,00$ | $99,75 \pm 0,07$ |

Les résultats obtenus sont suffisamment clairs pour ne pas avoir besoin de représenter graphiquement l'évolution de la proportion de chaque classe dans la population des RABS à $n$ éléments de la classe *Ev*. On remarque tout d'abord que la classe *Up* est très faiblement représentée (parmi les 100 000 RABS de taille 7 simulés, un seul correspond à la classe *Up*). Ce résultat était attendu, d'après ce que l'on connaît sur le lien entre les modes d'itération et l'apparition et la disparition des cycles limites. Le fait que les RABS de la classe *Up* soient en nombre très faible, est une bonne nouvelle pour la modélisation des réseaux de régulation génétique par des RABS, comme nous l'avons vu précédemment.

Nous pouvons également observer que la classe de comportement dynamique *Down* est de moins en moins représentée, à mesure que la taille des RABS augmente. Pour les RABS de taille 6 et 7, la quasi-totalité des réseaux de la classe *Ev* sont des des réseaux de la classe *None*. Ceci signifie donc que pour des RABS de "grande" taille, expliquer la disparition et l'apparition des cycles limites, sim-

plement par une (dé)synchronisation des éléments, ne suffit pas. Comme dans la partie précédente, nous pensons que c'est la connectivité très forte des RABS à $n$ éléments qui est à l'origine de ce phénomène. Nous n'avons pas réalisé de simulations permettant de le montrer. En revanche, nous pouvons imaginer que, pour des RABS de grande taille dont la connectivité est faible, nous pouvons avoir des modules de petite taille indépendants, comme pour le réseau de la figure 3.5. Dans un tel réseau, la dynamique globale du réseau dépend de la dynamique des modules de petite taille, et donc, le comportement dynamique est celui des réseaux de petite taille. Pour ces réseaux, la (dé)synchronisation des éléments explique en bonne partie l'apparition et la disparition des cycles limites. Nous verrons un réseau de grande taille de ce type dans la partie 3.3. Cependant, même pour les réseaux de petite taille, la proportion de la classe *None* est élevée. Ceci signifie donc que l'apparition et la disparition des cycles limites ne sont pas seulement liées à des (dé)synchronisations d'éléments. On observe par exemple, pour certains réseaux, que l'ordre des blocs des modes d'itération blocs-séquentiels peut également avoir de l'importance. Cette étude des RABS de la classe *None* n'est pas abordée ici.

La relation d'ordre que nous avons définie sur les modes d'itération blocs-séquentiels ne suffit pas, à elle seule, pour expliquer l'apparition et la disparition des cycles limites pour les RABS de la classe *Ev*. En revanche, pour tous les réseaux de cette classe, il existe des modes d'itération pour lesquels on observe aucun cycle limite. Il nous paraît intéressant de voir si le nombre de modes d'itération blocs-séquentiels avec (ou sans) cycles limites est important ou non. A partir des mêmes échantillons que pour les simulations précédentes, nous avons calculé le nombre moyen de modes d'itération pour lesquels on observe des cycles limites pour les réseaux de la classe *Ev*. Nous allons voir qu'il existe un lien entre cette proportion de modes avec cycles limites et les classes de comportement *Down*, *Up* et *None*. Afin de pouvoir comparer les résultats obtenus pour les différentes tailles de RABS, nous exprimons le nombre moyen de modes d'itération pour lesquels on observe des cycles limites sous la forme de proportion (exprimée en pourcentage) du nombre total de modes d'itération blocs-séquentiels. Le tableau 3.10 présente la proportion de modes d'itération blocs-séquentiels pour lesquels on observe des cycles limites pour les RABS à $n$ élements de la classe *Ev*, avec $2 \leq n \leq 7$. Ce tableau donne cette proportion pour chaque sous-classe de la classe de comportement dynamique *Ev*, ainsi que pour l'ensemble de cette classe.

On constate que, plus la taille des RABS est grande, plus la proportion des modes d'itération blocs-séquentiels avec cycles limites est grande, pour les RABS de la classe *Ev*. On passe environ du tiers des modes d'itération pour les réseaux de petite taille à un peu plus de la moitié des modes d'itération pour les réseaux de taille 7. L'évolution n'est pas la même pour les différentes sous-classes de la classe *Ev*. On observe par exemple que pour les RABS de la classe *Down*, plus leur taille est grande, plus rares sont les modes d'itération pour lesquels on observe des cycles limites. On peut expliquer ce phénomène par le fait que le diagramme de Hasse représentant la relation d'ordre entre les modes d'itération,

TAB. 3.10 – Proportion des modes d'itération pour lesquels on observe des cycles limites pour les RABS à $n$ éléments de la classe *Ev*, pour $2 \leq n \leq 7$.

| $n$ | *Down* (en %) | *Up* (en %) | *None* (en %) | Total (en %) |
|---|---|---|---|---|
| 2 | 33,33 | - | - | 33,33 |
| 3 | 26,52 | 82,34 | 40,47 | 31,96 |
| 4 | 22,45 | 83,94 | 38,93 | 34,27 |
| 5 | 19,08 | 86,43 | 41,72 | 39,56 |
| 6 | 16,36 | 92,21 | 48,23 | 47,48 |
| 7 | 13,04 | 98,86 | 57,26 | 57,10 |

est de plus en plus "large", au fur et à mesure que le nombre $n$ d'éléments du réseau augmente. La largeur du diagramme augmente plus rapidement que sa hauteur. La largeur maximale est de $n!$ pour la base (nombre de modes d'itération séquentiels), et sa hauteur est égale à $n$. Ainsi, si le fait de séparer 2 éléments dans les modes d'itération a pour effet de faire disparaître les cycles limites, la proportion de modes d'itération blocs-séqiuentiels concernés par cette séparation augmente avec la taille des réseaux, et donc la proportion de réseaux sans cycle limite augmente également. Pour les réseaux de la classe *None*, on observe des cycles limites pour de plus en plus de modes, au fur et à mesure que la taille des réseaux augmente. On pourrait dire que les RABS de la classe *Down* ont tendance à adopter un comportement similaire aux RABS de la classe *Fi* lorsque leur taille augmente, alors que les RABS de la classe *None* se rapprocheraient des RABS de la classe *Mi*. En ce qui concerne les rares réseaux de la classe *Up*, on constate que le nombre de modes d'itération pour lesquels on observe des cycles limites est très élevé, et augmente avec le nombre d'éléments des réseaux. Le comportement dynamique de ces RABS pourrait presque être confondu avec celui des réseaux de la classe *Mi*, à l'exception de rares modes d'itération, situés dans le haut du diagramme de Hasse, pour lesquels on n'observe pas de cycles limites.

Avant de clore cette partie concernant les RABS de la classe *Ev*, nous souhaitons montrer, sous la forme d'un exemple que l'on ne peut pas différencier les 3 classes *Down*, *Up* et *None*, simplement à partir des graphes d'interaction des réseaux correspondant à ces classes. On considère les 3 RABS à 3 éléments, $R^{Down} = (\mathbf{W}^{Down}, \boldsymbol{\theta}^{Down})$, $R^{Up} = (\mathbf{W}^{Up}, \boldsymbol{\theta}^{Up})$ et $R^{None} = (\mathbf{W}^{None}, \boldsymbol{\theta}^{None})$ définis comme suit :

$$\mathbf{W}^{Down} = \begin{pmatrix} 1 & -1 & 2 \\ -1 & -1 & 1 \\ 2 & 1 & -1 \end{pmatrix} \quad \boldsymbol{\theta}^{Down} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$$

$$\mathbf{W}^{Up} = \begin{pmatrix} 2 & -1 & 1 \\ -2 & -1 & 1 \\ 2 & 1 & -1 \end{pmatrix} \quad \boldsymbol{\theta}^{Up} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$$

$$\mathbf{W}^{None} = \begin{pmatrix} 2 & -1 & 1 \\ -2 & -2 & 1 \\ 2 & 1 & -1 \end{pmatrix} \quad \boldsymbol{\theta}^{None} = \begin{pmatrix} 1 \\ -3 \\ 1 \end{pmatrix}$$

Chacun de ces 3 réseaux appartient à l'une des sous-classes de *Ev* : $R^{Down}$ correspond à la classe de comportement *Down*, $R^{Up}$, à la classe de comportement *Up*, et $R^{None}$, à la classe de comportement *None*. Pourtant, ils partagent le même graphe d'interaction, présenté figure 3.7. La figure 3.8 montre les modes d'itération blocs-séquentiels pour lesquels chacun des réseaux a un cycle limite (les modes d'itération entourés correspondent aux modes avec cycles limites).

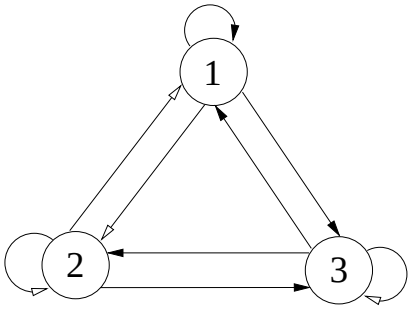



FIG. 3.7 – Graphe d'interaction commun aux réseaux $R^{Down}$, $R^{Up}$ et $R^{None}$.

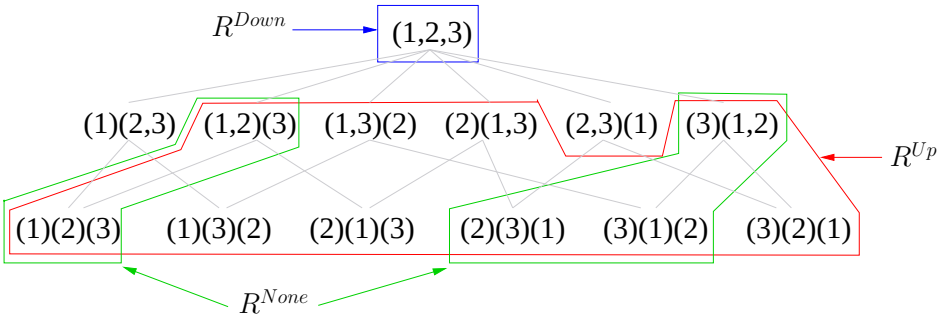



FIG. 3.8 – Modes d'itération blocs-séquentiels pour lesquels on observe des cycles limites pour le réseau $R^{Down}$ (en bleu), le réseau $R^{Up}$ (en rouge) et le réseau $R^{None}$ (en vert).

Le fait que l'on ne puisse pas réduire l'étude des classes de comportement dynamique à l'étude unique des graphes d'interaction n'est pas valable uniquement pour les réseaux de la classe *Ev*. Nous avons calculé que ce graphe d'interaction

est commun à 1 584 RABS de taille 3. Parmi ces 1584 réseaux, on trouve des réseaux appartenant à toutes les classes de comportement dynamique que nous avons définies. Les 188 968 RABS de taille 3 définissent en tout 1624 graphes d'interaction différents. Parmi ces graphes, seuls 473 correspondent à des réseaux qui appartiennent à une unique classe de comportement dynamique. Ainsi, la donnée du graphe d'interaction, c'est-à-dire l'ensemble des informations du type "qui agit sur qui, et de quelle façon (activation ou inhibition)" n'est pas suffisante, dans la majorité des cas, pour décrire un comportement dynamique général du réseau.

La relation d'ordre que nous avons définie sur l'ensemble des modes d'itération blocs-séquentiels n'a permis de décrire qu'une partie des apparitions et disparitions des cycles limites avec des changements de modes d'itération, pour les RABS de la classe *Ev*. Cette partie s'est, de plus, avérée être de moins en moins importante, au fur et à mesure que la taille des réseaux augmente. Nous avons également montré que, en moyenne, plus les RABS sont grands, plus il y a de modes d'itération pour lesquels on observe des cycles limites. Enfin, nous avons établi que l'on ne pouvait pas décrire exhaustivement les classes de comportement dynamique, en se basant uniquement sur les graphes d'interaction des réseaux. Dans la partie suivante, nous présentons un exemple de modèle de réseau de régulation génétique, sous la forme d'un RABS qui est de la classe *Down*.

## 3.3 L'exemple de la morphogénèse florale chez *Arabidopsis thaliana*

Nous nous intéressons dans cette partie à un exemple de modèle de réseau de régulation génétique, lié au processus de la morphogénèse florale chez *Arabidopsis thaliana*. Ce réseau est d'une importance majeure dans le cadre de notre travail. En effet, lors de notre master [89], nous avons commencé à en étudier la dynamique, et les résultats obtenus nous ont poussé à nous intéresser aux problèmes de robustesse dans les réseaux de régulation. Nous allons montrer comment nous pouvons utiliser des RABS de taille 3, dont nous avons simulé la dynamique dans la partie 3.1, afin de décrire la dynamique globale du système.

*Arabidopsis thaliana*, plante à fleurs, appelée communément "arabette des dames", est l'organisme modèle le plus utilisé en botanique, au même titre que la souris ou la drosophile pour les animaux. Son génome est entièrement séquencé et comporte environ 25 000 gènes. La morphogénèse florale décrit le processus de développement de la fleur. Chez les plantes à fleurs, on trouve au bout des tiges ce que l'on appelle un méristème apical. C'est un ensemble de cellules indifférenciées, qui vont donner des tiges, des feuilles ou des fleurs. La morphogénèse florale décrit donc le processus qui permet de passer d'un méristème apical avec des cellules indifférenciées à une fleur dont les cellules sont différenciées selon 4 types de tissus : les sépales, les pétales, les étamines et les carpelles (parties constitutives des pistils). Chez *Arabidopsis thaliana*, ces 4 types de tissus floraux sont organisés

en couronnes concentriques comme illustré par la figure 3.9. Les 4 sépales se trouvent sur la couronne extérieure, puis, vers le centre, les 4 pétales, ensuite les 6 étamines et enfin les 2 carpelles qui se trouvent au centre de la fleur.

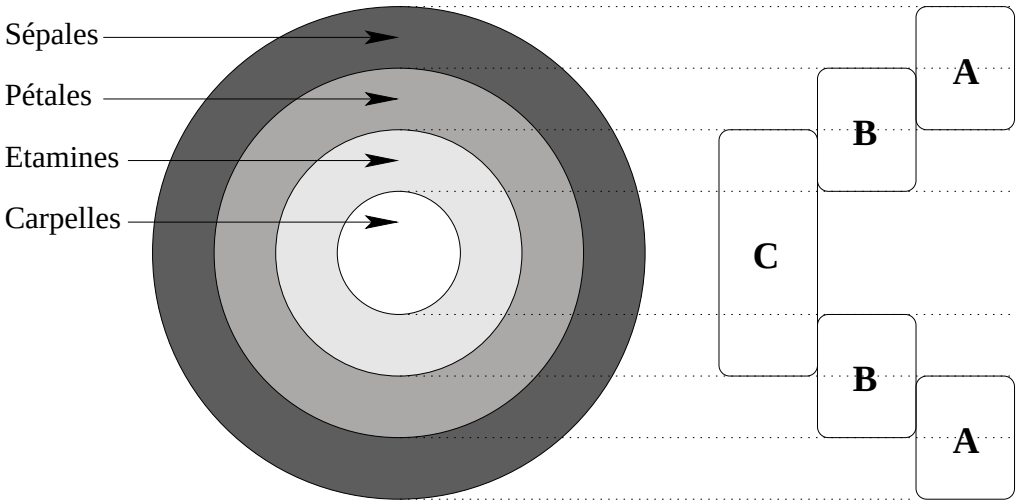



FIG. 3.9 – Schéma d'une coupe horizontale de la fleur d'*Arabidopsis thaliana* et correspondance avec le modèle ABC.

Coen et Meyerowitz [90] ont proposé un modèle afin de décrire la morphologie de la fleur d'*Arabidopsis thaliana*, appelé modèle ABC. D'après ce modèle il existe 3 activités pouvant être chacune présente dans 2 couronnes adjacentes de la fleur et dont la combinaison détermine la nature du tissu floral. L'activité A peut être présente dans les 2 couronnes extérieures, l'activité B dans les couronnes intermédiaires et l'activité C dans les 2 couronnes centrales (voir la figure 3.9). La présence de l'activité A seule, implique la formation des sépales. Si les activités A et B sont présentes ensemble, alors les pétales se forment. Les activités B et C forment ensemble les étamines. Enfin, la présence seule de l'activité C permet la formation des carpelles. Selon le modèle, les activités A et C s'inhibent mutuellement. Ainsi, lorsque l'une est présente, l'autre est absente, et inversement. A chaque activité correspond un ensemble de gènes donné. Si les gènes correspondant à l'activité A seuls sont actifs, alors les sépales se forment... Dans un modèle de réseau de régulation génétique lié à la morphogénèse florale chez *Arabidopsis thaliana*, on s'attend donc à ce que, pour chaque attracteur modélisant l'un des tissus floraux, les gènes correspondant à ce tissu dans le modèle ABC soient activés.

Mendoza et Alvarez-Buylla [44] ont proposé un modèle sous forme de réseau d'automates booléens à seuil pour le réseau de régulation génétique sous-jacent au processus de la morphogénèse florale chez *Arabidopsis thaliana*. Ce réseau est constitué de 12 automates correspondant à 11 gènes et à 1 complexe protéique en interaction avec ces gènes. Les 12 éléments sont les suivants : EMBRYONIC

FLOWER 1 (EMF1), TERMINAL FLOWER 1 (TFL1), LEAFY (LFY), APETALA1 (AP1), CAULIFLOWER (CAL), LEUNIG (LUG), UNUSUAL FLORAL ORGANS (UFO), BURST FORMING UNIT (BFU), AGAMOUS (AG), APETALA3 (AP3), PISTILLATA (PI) et SUPERMAN (SUP). Parmi ces gènes, AP1 est le seul correspondant à l'activité A du modèle ABC. BFU, AP3 et PI constituent l'activité B. AG est l'unique représentant de l'activité C. Se basant sur des éléments de la littérature relatifs aux gènes impliqués dans la développement floral, ils ont pu reconstruire le graphe d'interaction de ce réseau. Ce graphe d'interaction est présenté figure 3.10.

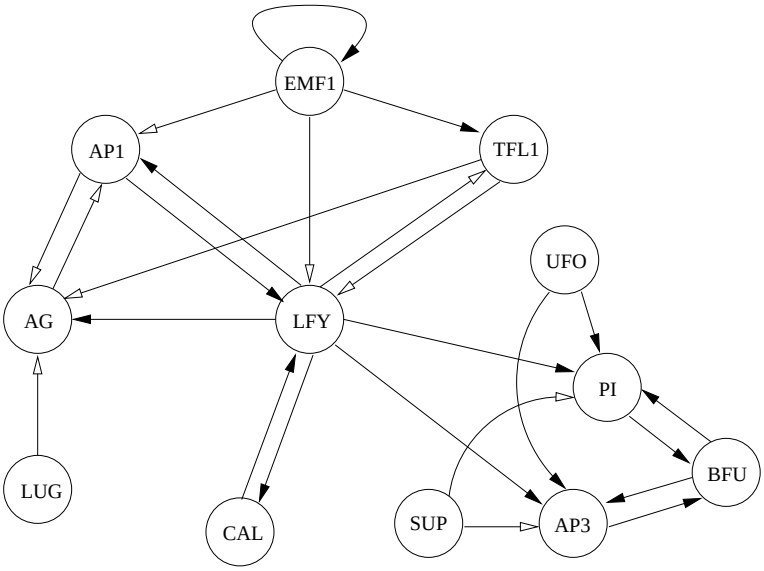



FIG. 3.10 – Graphe d'interaction du réseau de régulation de la morphogénèse florale chez *Arabidopsis thaliana*.

Afin de construire la matrice d'interaction et le vecteur seuil de leur RABS, Mendoza et Alvarez-Buylla ont également tiré de la littérature des inégalités entre les poids de la matrice, correspondant à des influences d'importance différente dans la régulation de certains gènes. Ils ont par exemple établi que l'activation de LFY par AP1 était plus importante que l'activation du même LFY par CAL. A partir de ces inégalités sur les poids, ils ont utilisé une procédure de reconstruction de la matrice d'interaction et du vecteur seuil, basée sur un algorithme génétique dont nous ne détaillerons pas le principe ici. Notons que, lors de la phase de construction de la matrice d'interaction et du vecteur seuil, Mendoza et Alvarez-Buylla n'ont pas utilisé de vecteurs d'interaction représentatifs de fonctions booléennes à seuil, comme nous les avons définis dans la partie 2.2. Nous allons voir l'importance de ceci dans le résultat obtenu. La matrice d'interaction $\mathbf{W}_{Men}$ et le vecteur seuil $\boldsymbol{\theta}_{Men}$ construits sont donnés ci-après (les gènes sont ordonnés comme lors de l'énumération précédente). Ils définissent le réseau que l'on note $R_{Men}$.

$$\mathbf{W}_{Men} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -2 & -1 & 0 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 5 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & -2 & 1 & -2 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 & 2 & 1 & 0 & 0 & 0 & -2 \\ 0 & 0 & 4 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad \boldsymbol{\theta}_{Men} = \begin{pmatrix} 0 \\ 0 \\ 3 \\ -1 \\ 1 \\ 0 \\ 0 \\ 1 \\ -1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Toute ligne de la matrice d'interaction contient au moins 8 poids nuls, c'est-à-dire que chaque gène est régulé par, au maximum, 4 éléments du réseau. Nous pouvons donc utiliser les vecteurs représentatifs des fonctions booléennes à seuil à 4 arguments donnés dans le tableau 2.4 et les tableaux 2 et 3 de l'annexe pour représenter les fonctions de transition locale du réseau $R_{Men} = (\mathbf{W}_{Men}, \boldsymbol{\theta}_{Men})$. Rappelons que ces vecteurs représentatifs nous permettent de visualiser les interactions réelles entre éléments du réseau. En notant $\mathbf{V}_k$ le vecteur d'interaction correspondant au gène $k$, on peut alors faire les remplacements équivalents suivants :

$$\begin{array}{lcl} \mathbf{V}_{TFL1} = (1,0,-2,0,0,0,0,0,0,0,0,0)(0) & \Longleftrightarrow & (1,0,-1,0,0,0,0,0,0,0,0,0)(0) \\ \mathbf{V}_{LFY} = (-2,-1,0,2,1,0,0,0,0,0,0,0)(3) & \Longleftrightarrow & (0,0,0,0,0,0,0,0,0,0,0,0)(0) \\ \mathbf{V}_{AP1} = (-1,0,5,0,0,0,0,0,-1,0,0,0)(-1) & \Longleftrightarrow & (-1,0,2,0,0,0,0,0,-1,0,0,0)(-1) \\ \mathbf{V}_{CAL} = (0,0,2,0,0,0,0,0,0,0,0,0)(1) & \Longleftrightarrow & (0,0,1,0,0,0,0,0,0,0,0,0)(0) \\ \mathbf{V}_{AP3} = (0,0,3,0,0,0,2,1,0,0,0,-2)(0) & \Longleftrightarrow & (0,0,2,0,0,0,1,1,0,0,0,-1)(0) \\ \mathbf{V}_{PI} = (0,0,4,0,0,0,1,1,0,0,0,-1)(0) & \Longleftrightarrow & (0,0,2,0,0,0,1,1,0,0,0,-1)(0) \end{array}$$

On remarque immédiatement que le gène LFY qui semble être régulé au travers de 2 activations et de 2 inhibitions, d'après le graphe d'interaction, n'est en fait pas régulé. Nous tenons à insister sur le fait que nous ne transformons pas le réseau. Chaque fonction de transition locale reste inchangée. Nous mettons cependant en évidence les poids qui correspondent à de véritables interactions. Dans le cas du gène LFY, les poids choisis par Mendoza et Alvarez-Buylla ne permettent pas d'avoir une somme totale d'activation plus élevée que le seuil. Le gène LFY ne peut donc jamais être activé. Tous les arcs du graphe d'interaction de la figure 3.10, qui ont pour destination le noeud LFY, ne sont pas pris en compte dans le RABS défini par la matrice d'interaction et le vecteur seuil proposés par Mendoza et Alvarez-Buylla. On peut également remarquer que les poids d'interaction situés à la $3^e$ coordonnée de chacun des vecteurs $\mathbf{V}_{AP1}, \mathbf{V}_{CAL}, \mathbf{V}_{AP3}$ et $\mathbf{V}_{PI}$ ont vu leur valeur diminuer dans les vecteurs d'interaction minimaux. Ces poids représentent l'influence du gène LFY dans la régulation des 4 gènes AP1,

CAL, AP3, PI. Mendoza et Alvarez-Buylla avaient imposé un ordre entre ces poids, pour modéliser le fait que LFY joue un rôle plus important dans l'activation de AP1 que dans celle de AP3 par exemple. Or, nous avons vu que lorsque l'on modélise les réseaux de régulation génétique par des RABS, comparer l'amplitude des poids d'une même colonne (qui représentent les différentes influences d'un même gène sur différents gènes) n'a pas de sens. On voit ici que c'est bien le cas, puisque, dans les vecteurs d'interaction minimaux, le $3^e$ poids de $\mathbf{V}_{AP1}$ est égal au $3^e$ poids de $\mathbf{V}_{AP3}$. En utilisant les vecteurs d'interaction minimaux, on obtient la matrice d'interaction $\mathbf{W'}$ et le vecteur seuil $\boldsymbol{\theta'}$ (on a $R_{Men} = (\mathbf{W'}, \boldsymbol{\theta'})$) suivants (nous rappelons l'ordre des éléments : EMF1, TFL1, LFY, AP1, CAL, LUG, UFO, BFU, AG, AP3, PI et SUP) :

$$\mathbf{W'} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & -2 & 1 & -2 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 2 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad \boldsymbol{\theta'} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ -1 \\ 0 \\ 0 \\ 0 \\ 1 \\ -1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

A partir de cette matrice d'interaction, on peut construire le graphe d'interaction de la figure 3.11. Ce graphe représente les réelles interactions entre les éléments du réseau. Les noeuds sont réorganisés afin de mettre en évidence certaines propriétés utiles pour la suite.

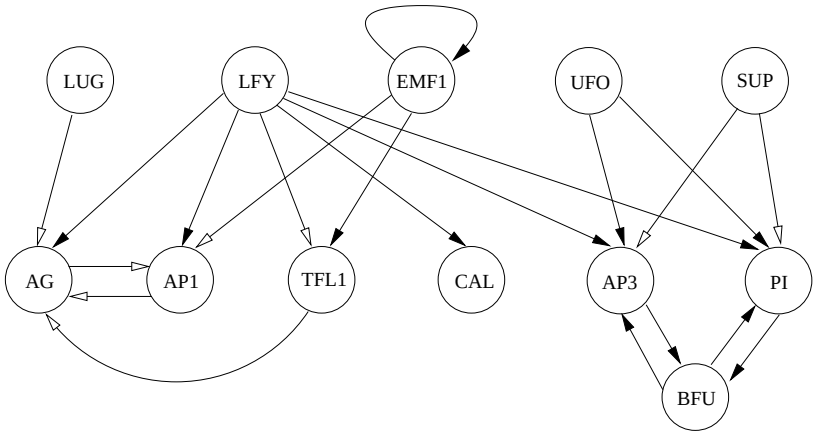



FIG. 3.11 – Graphe d'interaction correspondant à $\mathbf{W'}$ et $\boldsymbol{\theta'}$.

Intéressons-nous maintenant à la dynamique du système, et plus précisément aux attracteurs. Les remarques générales qui suivent sont vraies quel que soit le mode d'itération choisi.

On remarque, sur le graphe d'interaction donné figure 3.11, que les sommets $LUG$, $LFY$, $UFO$ et $SUP$ sont des sommets sources. Ils représentent des gènes qui ne sont pas régulés par les éléments du réseau. Quel que soit l'état initial de chacun des automates correspondant à ces sommets, l'état atteint après une itération restera donc constant pour les itérations suivantes. En notant $x_k(t)$ l'état de l'automate $k$ au temps $t$ et en considérant que $t = 0$ correspond au temps initial, on a même (d'après les fonctions de transition locale) :

$$x_{LUG}(t) = x_{LFY}(t) = x_{UFO}(t) = x_{SUP}(t) = 0, \text{ pour } t \geq 1.$$

Le gène CAL est régulé uniquement par le gène LFY et d'après la fonction de transition locale qui lui est associée, on a également :

$$x_{CAL}(t) = x_{LFY}(t-1) = 0, \text{ pour } t \geq 2.$$

On voit donc que les états des automates $LUG$, $LFY$, $UFO$, $SUP$ et $CAL$, dans les attracteurs, sont fixés à 0, quel que soit l'état initial du système. Les attracteurs du réseau ne dépendent alors que des états des automates $EMF1$, $TFL1$, $AG$, $AP1$, $AP3$, $PI$ et $BFU$.

Concernant les automates $EMF1$ et $TFL1$, on peut remarquer, toujours à partir de fonctions des transition locales définies par la matrice d'interaction et le vecteur seuil, que :

$$x_{EMF1}(t) = x_{EMF1}(0), \text{ pour t} \geq 0$$

et

$$x_{TFL1}(t) = x_{EMF1}(t-1) = x_{EMF1}(0), \text{ pour } t \geq 2$$

Pour $t \geq 2$, l'état de l'automate TFL1 est donc constant, égal à l'état de l'automate EMF1.

D'après ces remarques, on peut donc déduire les attracteurs du système entier à partir des attracteurs de 2 réseaux de taille 3 fortement connexes, notés $R_1$ et $R_2$. L'étude du comportement dynamique du RABS de 12 éléments peut donc se ramener à l'étude de 2 RABS de taille 3. Le réseau $R_1$ est composé des 3 éléments, $ET1$ (l'état de cet automate représente l'état des automates $EMF1$ et $TFL1$ pour $t \geq 2$), $AP1$ et $AG$. Le réseau $R_2$ est également un RABS à 3 éléments, $BFU$, $AP3$ et $PI$. On note $\mathbf{W}_1$ (resp. $\mathbf{W}_2$), la matrice d'interaction et $\boldsymbol{\theta}_1$ (resp. $\boldsymbol{\theta}_2$) le vecteur seuil qui correspondent au réseau $R_1$ (resp. $R_2$). Les poids d'interaction et les seuils sont obtenus à partir des éléments de $\mathbf{W}'$ et $\boldsymbol{\theta}'$, et en utilisant des vecteurs d'interaction représentatifs des fonctions booléennes à seuil définies par $\mathbf{W}'$ et $\boldsymbol{\theta}'$. Matrices d'interaction et vecteur seuils ainsi obtenus sont donnés ci-dessous (l'ordre des automates est celui donné dans la définition

de $R_1$ et de $R_2$).

$$\mathbf{W}_1 = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 0 & -1 \\ -1 & -1 & 0 \end{pmatrix} \qquad \boldsymbol{\theta}_1 = \begin{pmatrix} 0 \\ -1 \\ -1 \end{pmatrix}$$

$$\mathbf{W}_2 = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \qquad \boldsymbol{\theta}_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

De ces matrices d'interaction, on déduit les graphes d'interaction présentés figure 3.12.




FIG. 3.12 – Graphes d'interaction correspondant à $R_1$ à gauche et à $R_2$ à droite.

Les 2 réseaux $R_1$ et $R_2$ sont 2 RABS à 3 éléments. Nous avons réalisé la simulation de l'ensemble de tous les RABS de taille 3 dans la partie 3.1. Nous connaissons donc le comportement dynamique de ces 2 réseaux. $R_1$, comme $R_2$, est un réseau de la classe de comportement *Ev*, de la sous-classe de comportement *Down*.

Pour le réseau $R_1$, on a les 3 points fixes suivants : 100, 010 et 001. Si les automates $AG$ et $AP1$ sont dans un même bloc du mode d'itération, on a également le cycle limite de longueur 2 suivant : [011 , 000].

Pour le réseau $R_2$, on a les 2 points fixes suivants : 111 et 000. Pour le mode d'itération parallèle, on a également le cycle limite de longueur 2 suivant : [100 , 011].

Remarquons que Sené et Goles ont également étudié la dynamique du réseau de régulation génétique de la morphogénèse florale chez *Arabidopsis thaliana*, de façon similaire [25, 91]. En utilisant la symétrie de $R_2$ et la symétrie de la composante fortement connexe à 2 éléments de $R_1$, ils ont pu montrer que les cycles limites de ces 2 sous-réseaux sont de longueur 2, à partir d'un résultat précédent de Goles et Olivos [74].

Dans le réseau complet, les sous-réseaux $R_1$ et $R_2$ sont reliés par le gène LFY, qui régule les éléments de chacun des 2 sous-réseaux. Or, nous avons vu que, pour les attracteurs, l'état de LFY est 0. Ainsi, dans le réseau complet, si l'on ne s'intéresse qu'aux attracteurs, l'état du réseau $R_1$ est indépendant de l'état du réseau $R_2$. On peut donc immédiatement déduire des attracteurs de

$R_1$ et $R_2$, les attracteurs du réseau de Mendoza, $R_{Men}$, en associant à tous les attracteurs de $R_1$, l'ensemble des attracteurs de $R_2$. On peut également déduire du fait que $R_1$ et $R_2$ sont des réseaux de la classe de comportement *Down*, que le réseau $R_{Men}$ appartient également à *Down*. Nous avons donc un exemple de réseau de grande taille, dont la connectivité est faible, qui montre que la classe de comportement *Down* reste tout de même importante au fur et à mesure que le nombre d"éléments augmente.
Quel que soit le mode d'itération choisi, on observe 6 points fixes pour le réseau $R_{Men}$ :

- 110000010110
- 110000000000
- 000100010110
- 000100000000
- 000000001000
- 000000011110

En fonction des caractéristiques des blocs du mode d'itération choisi, on peut également décrire les cycles limites que le réseau atteint. Ces cycles sont tous de longueur 2 :

- si $AG$ et $AP1$ sont dans un même bloc du mode d'itération, et que $BFU$, $AP3$ et $PI$ sont également réunis au sein d'un même bloc (cas du mode parallèle, par exemple), alors on observe les 7 cycles limites suivants :
  - [110000010000 , 110000000110]
  - [000100010000 , 000100000110]
  - [000000011000 , 000000001110]
  - [000100011110 , 000000010110]
  - [000100001000 , 000000000000]
  - [000100011000 , 000000000110]
  - [000100001110 , 000000010000]
- si $AG$ et $AP1$ sont dans un même bloc du mode d'itération, et que $BFU$, $AP3$ et $PI$ ne sont pas tous dans un même bloc, alors on observe les 2 cycles limites suivants :
  - [000100011110 , 000000010110]
  - [000100001000 , 000000000000]
- si $AG$ et $AP1$ sont dans deux blocs distincts du mode d'itération, et que $BFU$, $AP3$ et $PI$ sont également réunis au sein d'un même bloc, alors, on observe les 3 cycles limites suivants :
  - [110000010000 , 110000000110]
  - [000100010000 , 000100000110]
  - [000000011000 , 000000001110]
- enfin, si $AG$ et $AP1$ sont dans deux blocs distincts du mode d'itération, et que $BFU$, $AP3$ et $PI$ ne sont pas tous dans un même bloc (cas de tous les modes séquentiels, par exemple), alors on n'observe aucun cycle limite.

Nous retrouvons bien les 6 points fixes présentés par Mendoza et Alvarez-Buylla. Dans leur article, ils donnent une justification biologique au fait d'itérer le réseau avec le mode d'itération suivant :

(EMF1,TFL1)(LFY,AP1,CAL)(LUG,UFO,BFU)(AG,AP3,PI)(SUP). Ces gènes sont regroupés en fonction de l'ordre dans lequel ils sont activés dans le processus de différenciation des cellules florales. Pour ce mode d'itération, on n'observe aucun cycle limite, puisque les gènes AG et AP3 sont bien dans 2 blocs différents, alors que BFU est séparé de AP3 et PI. Les seuls attracteurs que l'on obtient pour ce mode d'itération sont donc les 6 points fixes, pour lesquels on peut donner une signification biologique. Dans le point fixe 000100000000, seul le gène AP1 est actif. L'activité A est donc présente seule, et d'après le modèle ABC, on peut associer ce point fixe aux cellules du tissu de type sépale. Dans le point fixe 000100010110, les gènes AP1, AP3 et PI, ainsi que le complexe protéique BFU sont actifs. On a donc l'activité A et l'activité B qui sont présentes. D'après le modèle ABC, on peut associer ce point fixe aux cellules du tissu de type pétale. De même, on peut associer le point fixe 000000001000 au type cellulaire "carpelle", et le point fixe 000000011110 au type cellulaire "étamine". Les 2 points fixes restant 110000000000 et 110000010110 ne peuvent être associés à aucun type cellulaire floral, puisque les gènes EMF1 et TFL1 y sont actifs. Or, ces gènes sont des gènes inhibiteurs de toute activité florale. Le premier point fixe peut être associé aux cellules du méristème apical qui ne deviennent pas des cellules différenciées florales. Le deuxième point fixe ne correspond à aucun type cellulaire observé sur les spécimens sauvages d'*Arabidopsis thaliana*, mais Mendoza et Alvarez-Buylla affirment que l'on peut obtenir ce type cellulaire sur des mutants. Ces 6 points fixes valident le modèle de réseau de régulation sous forme de RABS, puisqu'il est en adéquation avec le modèle ABC. En revanche, les cycles limites que l'on peut observer pour différents modes d'itération n'ont aucune signification biologique. Cela signifie donc que le synchronisme dans le processus d'expression des différents gènes du réseau est très important, pour que le réseau conserve sa fonction.

Nous avons montré qu'en ré-écrivant la matrice d'interaction et le vecteur seuil de ce réseau de 12 éléments, à l'aide de vecteurs d'interaction minimaux, nous pouvions trouver une simplification, qui a permis de nous ramener à l'étude de 2 sous-réseaux de taille 3. Cette simplification montre que la régulation de LFY par les autres éléments du réseau n'est pas nécessaire pour décrire la morphogénèse florale chez *Arabidopsis thaliana*. Nous pouvons alors nous demander si, en conservant la régulation de LFY, on peut obtenir des attracteurs en accord avec le modèle ABC.

Pour répondre à cette question, nous allons garder la matrice d'interaction et le vecteur seuil sous la forme de $\mathbf{W}'$ et $\boldsymbol{\theta}'$, tout en complétant la ligne correspondant à la régulation de LFY (il s'agit de la $3^e$ ligne). D'après les éléments de la littérature retenus par Mendoza et Alvarez-Buylla, le rôle activateur de AP1 sur LFY est plus important que celui de CAL. Nous devons donc compléter la ligne 3 de $(\mathbf{W}', \boldsymbol{\theta}')$ par un vecteur d'interaction minimal correspondant à une régulation sous la forme de 2 activations d'importances différentes (représentés par 2 poids de valeurs positives différentes) et de 2 inhibitions. D'après le tableau 2 de l'annexe, il y a 21 vecteurs représentatifs de classes d'équivalence de fonctions

booléennes à seuil qui correspondent à ce cas. En prenant en compte le fait que les poids correspondant aux 2 inhibitions peuvent être permutés dans le cas où ils sont différents (nous n'avons aucune information sur l'importance relative de ces 2 inhibitions), nous obtenons 33 vecteurs d'interaction minimaux à partir desquels on peut reconstruire la ligne correspondant à la régulation de LFY. Nous avons simulé la dynamique des 33 RABS à 12 éléments correspondant. Comme nous nous intéressons aux points fixes de ces réseaux, afin de voir s'ils décrivent un comportement cohérent avec le modèle ABC, nous itérons les réseaux avec le mode parallèle uniquement. Parmi les 33 RABS simulés, aucun ne présente un comportement dynamique cohérent avec le modèle ABC. En particulier, aucun des 33 RABS simulés n'a de point fixe correspondant aux sépales. Ceci ne signifie pas que le fait de prendre en compte la régulation de LFY dans le réseau, empêche de décrire un comportement qui explique la morphogénèse florale. En effet, nous n'avons modifié aucun des poids des interactions ne concernant pas LFY. Il existe peut-être un réseau dont la matrice d'interaction correspond au graphe d'interaction 3.10 et qui permet d'expliquer la morphogénèse florale. Pour trouver ce réseau, il faudrait tester toutes les matrices d'interaction et tous les vecteurs seuils possibles correspondant à ce graphe d'interaction. Ce n'est pas le propos de notre étude.

En revanche, nous remarquons que, si l'on relâche la contrainte stipulant que que le rôle activateur de AP1 sur LFY est plus important que celui de CAL, on trouve 2 RABS qui sont en accord avec le modèle ABC. Pour l'un de ces 2 réseaux, AP1 et CAL sont d'importance égale dans la régulation de LFY ; pour l'autre, CAL a un rôle plus important que AP1. Chacun de ces 2 réseaux est obtenu en remplaçant la ligne 3 du système $(\mathbf{W}', \boldsymbol{\theta}')$, par l'un des 2 vecteurs minimaux suivants :
$\mathbf{V}^1_{LFY} = (-1, -1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0)(1)$,
$\mathbf{V}^2_{LFY} = (-1, -2, 0, 1, 2, 0, 0, 0, 0, 0, 0, 0, 0)(1)$.
Pour ces 2 réseaux, on observe 7 points fixes. 6 d'entre eux sont les mêmes 6 points fixes que ceux observés pour le réseau lorsque LFY n'est pas régulé. Le $7^e$ est le point fixe suivant : 001110010110. Dans ce point fixe, les gènes associés aux activités A et B du modèle ABC sont actifs (comme pour le point fixe associé aux pétales), ainsi que les gènes LFY et CAL. Nous pensons que l'on peut également associer ce point fixe aux pétales. En effet, d'après [92], CAL et AP1 sont 2 gènes redondants, issus d'un récent événement de duplication. On ne trouve cette duplication que pour les plantes de la famille *Brassicacae* à laquelle appartient *Arabidopsis thaliana*. De plus, toujours d'après [92], LFY est activé et active l'expression de CAL et AP1, au tout début de la différientiation des cellules dans les méristèmes apicaux. On trouve AP1 exprimé très peu de temps après avoir observé que LFY est exprimé. On peut donc penser que le point fixe 001110010110 correspond à des pétales pour lesquels LFY et CAL sont restés activés. Nous avons donc trouvé 2 réseaux qui décrivent la morphogénèse florale chez *Arabidopsis thaliana*, tout en permettant de considérer que LFY est régulé dans le réseau.

Dans cette partie, nous avons montré que, pour l'étude de modèles de réseau de régulation génétique, nous pouvions parfois simplifier les réseaux de manière à décrire le comportement dynamique général du réseau, à partir du comportement de sous-réseaux spécifiques. Ceci montre l'intérêt d'étudier les RABS de petite taille, que l'on considère comme sous-modules de plus grands réseaux. Nous avons également montré l'importance de travailler avec des vecteurs d'interaction minimaux, pour définir les fonctions de transition locale des réseaux d'automates booléens à seuil.

## 3.4 *miARN* et robustesse

Nous avons décrit jusqu'à présent la régulation génétique par les seules interactions entre gènes et produits synthétisés à partir de ces gènes. Dans cette partie, nous allons aborder un autre type de régulation, liée à des ARN, que l'on appelle micro-ARN, et que l'on note *miARN*. Ces *miARN* sont de tout petits ARN simple brin contenant, après maturation à partir d'un forme "hair-pin", une vingtaine de nucléotides. Ils interviennent dans la régulation génétique en s'appariant aux ARN messagers. Ils ont un rôle de répresseurs et dégradent les ARN auxquels ils s'apparient (seuls ou avec des peptides auxquels ils sont complexés) et répriment leur traduction en protéine. Leur rôle exact et leur importance ne sont pas encore bien compris, mais quelques études montrent qu'ils interviennent dans la régulation de réseaux liés à un grand nombre de fonctions physiologiques essentielles [93, 94, 95]. Notons que certains gènes, comme celui qui exprime la protéine *p53*, contrôlent la transcription de certains *miARN*, ce qui peut introduire des circuits supplémentaires dans les réseaux de régulation, ces *miARN* n'étant plus de simples éléments sources des réseaux [47, 96, 97].

Notre propos n'est pas d'étudier de façon exhaustive le rôle de ces *miARN* dans la régulation génétique, mais nous proposons un protocole de simulation très simple, afin d'entrevoir quel peut être l'effet de ces répresseurs sur la robustesse des réseaux de régulation. Pour cela, nous étudions l'effet d'un élément répresseur sur un RABS de taille 3, dont le graphe d'interaction est fortement connexe. Nous choisissons de travailler uniquement sur des réseaux au graphe d'interaction fortement connexe, car ils nous semblent être les plus représentatifs des modules dont sont constitués les graphes d'interaction des réseaux de régulation génétique réels. Pour tous les RABS de taille 3, dont le graphe d'interaction est fortement connexe, nous suivons le protocole suivant :

1. nous simulons le comportement dynamique du réseau sans *miARN* ;
2. nous modélisons l'effet inhibiteur qu'aurait un *miARN* sur l'élément 1 du réseau en ajoutant 1 à la valeur du seuil correspondant à l'élément 1 ; nous simulons le comportement dynamique du réseau ainsi transformé ;
3. nous répétons l'étape précédente pour les éléments 2 et 3 du réseau.

Ainsi, à partir de chaque RABS de taille 3, on construit 3 nouveaux réseaux dont on simule le comportement dynamique.

Comme nous proposons d'étudier le rôle qu'aurait un *miARN* sur la robustesse des réseaux, nous notons, pour chaque RABS simulé, la classe de comportement dynamique à laquelle il appartient. Nous avons réalisé nos simulations à partir des 162 780 RABS de taille 3, dont le graphe d'interaction est fortement connexe. Le tableau 3.11 donne la proportion de chacune des classes de comportement dynamique dans la population des RABS étudiés, avant ajout du *miARN*. Nous pouvons déjà noter que la proportion de réseaux de la classe de

TAB. 3.11 – Répartition des RABS de taille 3 dont le graphe est fortement connexe.

| *Cy* | *Fi* | *Mi* | *Ev* | Total |
|---|---|---|---|---|
| 60846 | 49071 | 19457 | 33406 | 162780 |
| $37,4\%$ | $30,1\%$ | $12,0\%$ | $20,5\%$ | 100% |

comportement *Ev* est plus importante pour les RABS dont le graphe d'interaction est fortement connexe que pour la population totale, pour laquelle elle est de $18,49\%$.

A partir de chacun des réseaux de taille 3, nous avons construit, puis simulé, 3 nouveaux réseaux. Nous avons donc 488 340 nouveaux réseaux, dont la répartition en fonction des classes de comportement est donnée dans le tableau 3.12.

TAB. 3.12 – Répartition des réseaux en fonction des classes de comportement dynamique, après ajout de l'effet du *miARN*.

| *Cy* | *Fi* | *Mi* | *Ev* | Total |
|---|---|---|---|---|
| 168925 | 192673 | 49959 | 76783 | 488340 |
| $34,6\%$ | $39,5\%$ | $10,2\%$ | $15,7\%$ | 100% |

On remarque que la proportion de réseaux de la classe *Fi* a assez fortement augmenté avec l'ajout du *miARN*. Inversement, la classe de comportement *Ev* a vu sa proportion diminuer de façon significative. En terme de robustesse, on peut donc dire que l'ajout du *miARN* a tendance à rendre les modules de taille 3 plus robustes aux changements de modes d'itération.

Nous pouvons étudier l'influence du *miARN* de manière plus fine, en observant la distribution des RABS après prise en compte de l'effet du *miARN* en fonction de leur classe de comportement et de la classe de comportement du réseau de taille 3 à partir duquel ils sont construits. Cette répartition peut nous indiquer par exemple en quel type de réseau un réseau de la classe *Ev* est préférentiellement transformé. Le tableau 3.13 donne les distributions conditionnelles

de la classe de comportement, selon la classe de comportement du RABS de taille 3 initial.

TAB. 3.13 – Distributions conditionnelles de la classe de comportement (en colonnes), selon la classe de comportement du réseau initial (en lignes).

| | *Cy* | *Fi* | *Mi* | *Ev* | Total |
|---|---|---|---|---|---|
| *Cy* | 131091 | 23363 | 7783 | 20301 | 182538 |
| | 71,8% | 12,8% | 4,3% | 11,1% | 100% |
| *Fi* | 9765 | 117688 | 11003 | 8757 | 147213 |
| | 6,6% | 79,9% | 7,5% | 5,9% | 100% |
| *Mi* | 10684 | 18127 | 22176 | 7384 | 58371 |
| | 18,3% | 31,1% | 38,0% | 12,7% | 100% |
| *Ev* | 17385 | 33495 | 8997 | 40341 | 100218 |
| | 17,3% | 33,4% | 9,0% | 40,3% | 100% |

D'après les distributions conditionnelles, nous comprenons la raison pour laquelle la classe *Fi* est relativement plus importante après ajout de l'effet du *miARN* : environ 1/3 des réseaux des classes *Mi* et *Ev* ont été transformés en réseaux de la classe *Fi*. Cela signifie que l'ajout de l'inhibition par le noeud supplémentaire a tendance à supprimer les cycles limites des réseaux des classes *Mi* et *Ev*. De plus, la classe *Fi* est celle qui est la moins sensible à l'ajout de l'inhibition, puisque seulement 20% des réseaux de cette classe ont été transformés en réseaux de classe différente. Ensuite, vient la classe *Cy*, pour laquelle environ 30% des réseaux changent de classe de comportement dynamique avec l'ajout du *miARN*. Les classes avec un seul type d'attracteurs (soit des points fixes pour la classe *Fi*, soit des cycles limites pour la classe *Cy*) sont donc les moins sensibles à l'ajout du *miARN* inhibiteur.

Ceci confirme notre conclusion intermédiaire : l'ajout du *miARN* a tendance à rendre les réseaux plus robustes aux changements de modes d'itération, notamment en supprimant les cycles limites des réseaux avec points fixes et cycles limites. D'un point de vue biologique, on peut penser que les cycles limites ne sont, en grande partie, pas souhaitables pour de tels réseaux, comme par exemple celui lié à la morphogénèse florale chez *Arabidopsis thaliana*. Les *miARN* agiraient donc comme des "garde-fous", en empêchant des comportements cycliques.

Nous pouvons clore cette étude très rapide, en observant la taille moyenne du plus petit bassin d'attraction pour les réseaux, avant ajout du *miARN* et après ajout. Si cette taille diminue de façon significative avec l'ajout du *miARN*, nous pouvons en conclure que celui-ci a pour effet de supprimer les attracteurs à petits bassins. D'un point de vue biologique, de tels attracteurs ne représentent pas des types cellulaires intéressants. Le tableau 3.14 donne la taille moyenne du plus petit bassin d'attraction, pour les réseaux avant ajout du *miARN*. Cette taille moyenne est donnée pour chaque classe de comportement, et pour la totalité

des réseaux. Le tableau 3.15 présente la taille moyenne du plus petit bassin d'attraction, pour les réseaux après ajout du *miARN*. Pour chaque réseau, la taille du plus petit bassin d'attraction est calculée pour tous les modes d'itération. C'est la moyenne sur les modes d'itération qui est retenue. Nous rappelons que, pour les réseaux de taille 3, il y a $2^3 = 8$ configurations possibles. La taille maximale des bassins d'attraction est donc de 8.

TAB. 3.14 – Taille moyenne du plus petit bassin d'attraction pour les réseaux avant ajout du *miARN*.

| *Cy* | *Fi* | *Mi* | *Ev* | Total |
|---|---|---|---|---|
| 7,38 | 4,25 | 2,33 | 4,88 | 5,32 |

TAB. 3.15 – Taille moyenne du plus petit bassin d'attraction pour les réseaux après ajout du *miARN*.

| *Cy* | *Fi* | *Mi* | *Ev* | Total |
|---|---|---|---|---|
| 7,53 | 4,94 | 2,83 | 5,14 | 5,61 |

Nous observons une augmentation de la taille du plus petit bassin d'attraction, avec l'ajout du *miARN* inhibiteur. Cependant, cette augmentation est plutôt faible. Ainsi, on ne peut conclure à un effet spécial du *miARN* sur les attracteurs à bassin d'attraction de petite taille, puisque cette augmentation de la taille moyenne du plus petit bassin peut s'expliquer par la diminution du nombre moyen d'attracteurs.

Les réseaux de la classe *Mi* sont ceux pour lesquels l'augmentation relative de la taille du plus petit bassin d'attraction est la plus prononcée. Nous remarquons d'ailleurs que, pour ces réseaux, la taille du plus petit bassin est très faible, comparativement aux autres. Nous avons vu, dans la partie 3.1.2, que les RABS de cette classe sont ceux qui ont le plus d'attracteurs (ils ont au moins un point fixe et un cycle limite par mode d'itération). Cependant, ceci ne suffit pas à expliquer la taille si faible du plus petit bassin d'attraction. Il semble donc que ces réseaux sont caractérisés, de façon générale, par un attracteur dont le bassin d'attraction est de petite taille. A l'opposé, on constate que, pour les réseaux de la classe *Cy*, la taille moyenne du plus petit bassin d'attraction est très élevée. Ceci nous indique donc que ces réseaux ont un nombre moyen de cycles limites proches de 1 (ce que nous avions vu dans la partie 3.1.2), mais également que, lorsqu'il y a plusieurs cycles limites, la taille des bassins d'attraction de ces cycles est équilibrée.

Nous avons montré, au moyen de simulations relativement aisées à mettre en oeuvre, que l'on pouvait donner quelques pistes sur l'effet d'un *miARN* sur la robustesse des réseaux. D'après les résultats obtenus, nous pouvons supposer que

les *miARN* ont pour rôle d'augmenter la robustesse des réseaux de régulation. Cette conclusion est en accord avec les études récentes qui décrivent notamment une dérégulation des *miARN*, dans le cas de cancers [96].

Ce chapitre a porté sur les résultats des différentes simulations que nous avons réalisées concernant les réseaux d'automates booléens. Nous avons commencé par étudier le comportement général des RABS à $n$ éléments, avec $n \leq 7$. Les résultats obtenus montrent que le fait d'augmenter la taille, mais, surtout, la connectivité des réseaux, a pour effet de faire diminuer la robustesse de ces réseaux face à des perturbations des modes d'itération. En faisant augmenter la taille des réseaux, tout en conservant une connectivité faible, nous avons vu que les réseaux pouvaient retrouver une certaine robustesse. Nous avons fait le lien avec les graphes d'interaction des réseaux de régulation génétique connus, qui sont composés de modules de petite taille et fortement connexes, reliés par des arcs et des noeuds intermédiaires de degrés faibles. Nous avons également étudié les attracteurs des RABS et montré que leur nombre, ainsi que la longueur des cycles limites, augmentaient moins fortement avec la taille des réseaux que ce que montrent les études sur les réseaux booléens aléatoires.

Nous avons ensuite focalisé notre travail sur les réseaux de la classe de comportement la moins robuste aux changements de mode d'itération. Nous avons tenté de donner une explication aux apparitions et disparitions des cycles limites, observées pour ces réseaux, lorsque l'on modifie le mode d'itération. Nous avons également montré que l'on ne pouvait pas décrire les différentes classes de comportement dynamique, en décrivant seulement les graphes d'interaction des réseaux.

Enfin, nous avons traité un modèle de réseau de régulation génétique, pour lequel nous avons prouvé toute l'importance d'utiliser des vecteurs d'interaction minimaux pour définir les fonctions booléennes à seuil. Nous avons montré comment l'on pouvait ramener l'étude du réseau de 12 gènes à la description de la dynamique de 2 RABS de taille 3. Nous avons clos ce chapitre, en présentant une étude très simple visant à étudier le rôle général que peuvent avoir les *miARN* vis-à-vis de la robustesse des réseaux de régulation génétique.

Nous terminons notre manuscrit par le prochain chapitre, qui présente quelques résultats théoriques sur les cycles limites dans les réseaux d'automates.

# Chapitre 4

# Itération bio-inspirée

Nous étudions la robustesse des réseaux d'automates booléens à seuil, dans le cadre de la modélisation des réseaux de régulation génétique. Dans ce chapitre, nous allons présenter une itération qui nous semble plus appropriée à la modélisation de ces réseaux biologiques que l'itération traditionnelle que nous avons utilisée jusqu'à ce point. A partir de la définition de cette itération que nous appelons bio-inspirée, nous pouvons énoncer quelques résultats concernant les cycles limites observés pour un réseau d'automates. Ce chapitre ne concerne plus la seule classe des réseaux d'automates booléens à seuil, mais celle des réseaux d'automates de façon générale. En revanche, pour conserver l'homogénéité du manuscrit, les exemples présentés concerneront uniquement des réseaux d'automates booléens à seuil.

## 4.1 Motivation et premières définitions

Nous avons défini, dans la partie 1.3.1, les modes d'itération blocs-séquentiels. Cette définition, qui correspond à ce que nous appelons l'itération traditionnelle des réseaux d'automates, stipule que, quelque soit le mode d'itération choisi, à chaque transition d'état du réseau, l'état de tous les éléments est mis à jour. Nous avons, en effet, déjà remarqué qu'entre deux temps consécutifs, $t$ et $t+1$, l'ensemble de toutes les fonctions de transition locale était appliqué. Ceci signifie, dans le cadre de la modélisation des réseaux de régulation génétique, qu'entre deux configurations consécutives du réseau, tous les gènes ont eu le temps de voir leur expression changer. Or, nous avons vu dans cette même partie 1.3.1 quelle était la signification biologique que l'on pouvait donner aux modes d'itération blocs-séquentiels. Si nous pouvons déterminer qu'il existe un intervalle de temps suffisamment important entre l'expression (au travers de tout le processus d'expression des gènes incluant le remodelage de la chromatine, l'action de l'ARN polymérase...) des gènes $i$ et $j$, alors les éléments $i$ et $j$ qui modélisent ces gènes sont placés dans deux blocs différents du mode d'itération blocs-séquentiel. Si l'on suppose que le gène $i$ est exprimé avant le gène $j$, alors, dans un tel cas, le gène $j$ peut être régulé par les produits synthétisés à partir du gène $i$ qui corres-

pondent à l'état actuel d'activation de $i$. On place donc $j$ dans un bloc suivant le bloc de $i$. S'il existe un intervalle de temps suffisamment long pour que le gène $j$ puisse dépendre de l'état actuel du gène $i$, alors il nous semble cohérent de considérer que l'on peut observer l'état du réseau entre le moment où le gène $i$ est exprimé et celui où le gène $j$ est exprimé. D'un point de vue expérimental, on étudie l'évolution temporelle des niveaux d'expression des différents gènes à l'aide de bio-puces du type microarrays ou puces à oligonucléotides [9, 10, 11, 12]. Il nous paraît improbable qu'entre 2 mesures de l'expression génétique par ces bio-puces, l'état de tous les gènes d'un même réseau ait eu le temps d'être mis à jour un même nombre de fois.

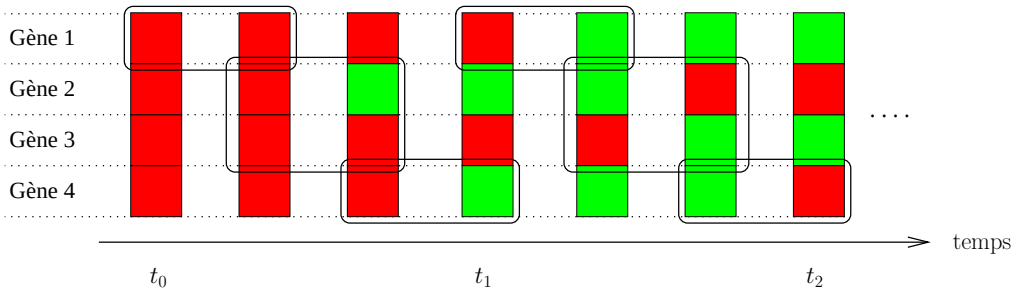



FIG. 4.1 – Schéma représentant l'évolution temporelle de l'expression de 4 gènes.

Le schéma simplifié de la figure 4.1 présente un exemple fictif d'évolution temporelle de l'expression de 4 gènes, à partir d'un état d'expression initial. A chaque gène correspond une case dont la couleur modélise l'état d'activation du gène : vert, le gène est exprimé, rouge, il n'est pas exprimé (initialement, aucun gène n'est donc exprimé). On suppose que les 4 gènes, notés de 1 à 4, sont regroupés de la façon suivante : $(1)(2,3)(4)$ ; c'est-à-dire que l'état d'expression du gène 1 est mis à jour, puis les états d'expression des gènes 2 et 3 sont mis à jour de façon synchrone, ensuite, l'état d'expression du gène 4 est mis à jour, et on réitère le processus à partir du gène 1. Sur le schéma, toute configuration obtenue à partir de la mise à jour d'un seul groupe est notée. Pour cet exemple, le mode d'itération blocs-séquentiel associé est, naturellement, $(1)(2,3)(4)$. Si l'on itère le RABS qui modélise le réseau fictif auquel sont associées les configurations du schéma, avec ce mode d'itération, on n'observe que les configurations correspondant aux temps $t_0$, $t_1$, $t_2$. En effet, on considère qu'entre 2 configurations successives du réseau, l'état de tous les éléments du réseau est mis à jour. Afin de modéliser le plus fidèlement, et le plus précisément possible, l'évolution de l'expression génétique au sein d'un réseau de régulation, nous décidons de définir une nouvelle façon d'itérer les réseaux d'automates qui nous permette de prendre en compte toutes les configurations représentées sur la figure 4.1.

A partir de ce constat, nous re-définissons les modes d'itération blocs-séquentiels

pour les réseaux d'automates. Nous appelons cette nouvelle façon d'itérer les réseaux, l'itération bio-inspirée. Dans l'itération traditionnelle, à chaque transition d'état du réseau, l'ensemble des éléments du réseau est mis à jour. Dans l'itération bio-inspirée, à chaque transition d'état du réseau, un groupe d'éléments (ceux regroupés dans un même bloc du mode d'itération blocs-séquentiel) du réseau est mis à jour. Nous définissons les modes d'itération blocs-séquentiels pour l'itération bio-inspirée à partir des partitions ordonnées.

Pour la suite, on considère que toute configuration initiale correspond au temps $t = 0$. On considère également que si $R = (Q, F)$ est un réseau d'automates à $n$ éléments et la fonction $f_i, i \in \{1, \cdots, n\}$, désigne la composante $i$ de $F$, c'est-à-dire, la fonction de transition locale de l'élément $i$ du réseau $R$.

On désigne l'ensemble des parties de $\{1, \cdots, n\}$ par $\mathcal{P}(\{1, \cdots, n\})$. Pour toute fonction de transition globale $F$ d'un réseau d'automates à $n$ éléments $R = (Q, F)$, nous définissons l'application $G_F : \mathcal{P}(\{1, \cdots, n\}) \times Q^n \to Q^n$, par :

$$\forall E \in \mathcal{P}(\{1, \cdots, n\}), \forall \mathbf{X} \in Q^n, \ \ G_F(E, \mathbf{X}) = \mathbf{Y}$$

avec $\forall i \in \{1, \cdots, n\}, \ y_i = \begin{cases} f_i(\mathbf{X}) & \text{si } i \in E \\ x_i & \text{sinon} \end{cases}$.

Cette application nous permet de donner la définition des modes d'itération blocs-séquentiels bio-inspirés suivante.

**Définition 4.1** (Mode d'itération blocs-séquentiel bio-inspiré)
*Soit $R = (Q, F)$ un réseau d'automates à $n$ éléments, et $P = (P_1, \cdots, P_p)$ une partition ordonnée de l'ensemble $\{1, \cdots, n\}$.*
*On dit que $R$ est itéré avec le* mode d'itération blocs-séquentiel bio-inspiré *(MIBSBI) associé à $P$ quand on a :*

$$\forall t \in \mathbb{N}, \ \ \mathbf{X}(t+1) = G_F(P_k, \mathbf{X}(t)), \quad \textit{avec } k \in \{1, \cdots, p\}, \ k \equiv t+1 \ \ [p].$$

Comme pour les modes d'itération blocs-séquentiels de l'itération traditionnelle, on peut définir les MIBSBI pour tout réseau d'automates, et pas uniquement pour les réseaux d'automates booléens à seuil.

Tout mode d'itération blocs-séquentiel de l'itération traditionnelle et tout MIBSBI étant définis à partir d'une partition ordonnée de $\{1, \cdots, n\}$, on peut faire correspondre à chaque MIBSBI un unique mode d'itération blocs-séquentiel de l'itération traditionnelle, et inversement. Ces modes sont associés à la même partition ordonnée. Nous verrons dans la suite comment les dynamiques obtenues à partir de deux modes d'itération correspondants sont liées. En particulier, nous pouvons définir les modes parallèle et séquentiels bio-inspirés, à partir, respectivement de la partition réduite à l'ensemble $\{1, \cdots, n\}$ et des partitions à $n$ blocs contenant chacun un unique élément (si $P = (P_1, \cdots, P_p)$ est une partition, alors on appelle tout $P_k$ un bloc de $P$). On peut remarquer immédiatement que les modes parallèles de l'itération traditionnelle et de l'itération bio-inspirée sont identiques. Dans les 2 cas, on a en effet (en conservant les notations de la définition 4.1) :

$$\forall t \in \mathbb{N}, \forall i \in \{1, \cdots, n\}, \ \ x_i(t+1) = f_i(\mathbf{X}(t)).$$

En revanche, les modes séquentiels sont bien différents, comme le montre l'exemple 4.1

*Exemple* 4.1 *:* On considère le réseau d'automates booléens à seuil à 3 éléments $R = (\mathbf{W}, \boldsymbol{\theta})$ dont la matrice d'interaction et le vecteur seuil sont donnés ci-après :

$$\mathbf{W} = \begin{pmatrix} 2 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \qquad \boldsymbol{\theta} = \begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix}$$

On part de la configuration initiale $\mathbf{X}(0) = 000$ et on itère le réseau avec le mode d'itération séquentiel associé à la partition ordonnée (3)(2)(1), dans le cas de l'itération traditionnelle, et dans le cas de l'itération bio-inspirée. On note les 2 premières configurations (autres que la configuration initiale) dans le cas de l'itération traditionnelle, et les 6 premières configurations dans le cas de l'itération bio-inspirée. On trouve alors les trajectoires suivantes :

- itération traditionnelle : $000 \to 011 \to 000$
- itération bio-inspirée : $000 \to 001 \to 011 \to 011 \to 010 \to 000 \to 000$

Les premières configurations de ces 2 trajectoires issues d'un même état global initial de $R$ sont différentes. Les modes séquentiels traditionnel et bio-inspiré ne sont donc pas identiques. On peut remarquer que la configuration 001 n'est pas présente dans la trajectoire correspondant à l'itération traditionnelle, alors que toute configuration de la trajectoire correspondant à l'itération traditionnelle est présente dans la trajectoire correspondant à l'itération bio-inspirée. On peut également remarquer que dans une trajectoire correspondant à l'itération bio-inspirée, on peut avoir deux configurations successives égales, sans avoir atteint de point fixe. A une unique configuration peuvent correspondre deux configurations suivantes différentes, selon le bloc qui est mis à jour durant la transition. Nous reviendrons plus loin sur cette remarque.

**Notation 4.1 :** *Soit $R$ un réseau d'automates à $n$ éléments, et $P$ une partition ordonnée de $\{1, \cdots, n\}$.*
*On note $(R, P, trad)$ le réseau $R$ itéré avec le mode blocs-séquentiel associé à $P$ pour l'itération traditionnelle, et $(R, P, bio)$ le réseau $R$ itéré avec le MIBSBI associé à $P$.*

A partir de la définition 4.1, pour tout réseau d'automates $R$, et toute partition ordonnée $P = (P_1, \cdots, P_p)$ de $\{1, \cdots, n\}$, on peut définir toute trajectoire issue d'une configuration initiale $\mathbf{X}(0)$, pour $(R, P, bio)$, comme la suite $(\mathbf{X}^k)_{k\in\mathbb{N}}$ définie par $\mathbf{X}^0 = \mathbf{X}(0)$ et $\forall k \in \mathbb{N},\ \mathbf{X}^{k+1} = G_F(P_{j_k}, \mathbf{X}^k)$, avec $j_k \in \{1, \cdots, p\},\ j_k \equiv k+1\ [p]$. Entre 2 configurations $\mathbf{X}^k$ et $\mathbf{X}^{k+p}$, $k \in \mathbb{N}$, l'état de tous les éléments de chaque bloc de $P$ est mis à jour, et donc chaque élément du réseau est mis à jour. Pour les modes d'itération blocs-séquentiels de l'itération traditionnelle, à chaque transition d'état du réseau, l'ensemble des éléments est mis à jour. On trouve donc la trajectoire issue de la même configuration initiale, $\mathbf{X}(0)$, pour $(R, P, trad)$, en prenant la suite extraite $(\mathbf{Y}^k)_{k\in\mathbb{N}}$ de $(\mathbf{X}^k)_{k\in\mathbb{N}}$, définie par $\forall k \in \mathbb{N},\ \mathbf{Y}^k = \mathbf{X}^{pk}$.

Dans l'exemple 4.1, les premiers états de la trajectoire correspondant à l'itération bio-inspirée sont $\mathbf{X}^0 = 000,\ \mathbf{X}^1 = 001,\ \mathbf{X}^2 = 011,\ \mathbf{X}^3 = 011,\ \mathbf{X}^4 = 010,\ \mathbf{X}^5 = 000,\ \mathbf{X}^6 = 000$. Pour l'itération traditionnelle, les premiers états de la trajectoire sont $\mathbf{Y}^0 = 000,\ \mathbf{Y}^1 = 011,\ \mathbf{Y}^2 = 000$. On retrouve bien $\mathbf{Y}^k = \mathbf{X}^{3k}$ pour $0 \leq k \leq 2$.

La définition des cycles limites que nous avons donnée dans la partie 1.2.1 ne peut pas s'appliquer à l'itération bio-inspirée. Nous avions en effet présenté tout cycle limite de longueur $l$ comme un ensemble de $l$ configurations deux à deux distinctes. Or, nous avons vu dans l'exemple précédent, que pour l'itération bio-inspirée, on peut avoir 2 configurations successives égales qui ne représentent pas un point fixe. Dans un cycle limite de l'itération bio-inspirée, on peut donc avoir deux configurations égales. Afin de définir les cycles limites pour l'itération bio-inspirée, nous utilisons le fait que pour tout réseau d'automates $R = (Q, F)$ à $n$ éléments, pour toute partition ordonnée $P = (P_1, \cdots, P_p)$ de $\{1, \cdots, n\}$, et pour toute configuration initiale $\mathbf{X}(0) \in Q^n$, la trajectoire issue de $\mathbf{X}(0)$, pour $(R, P, trad)$, est une suite extraite de la trajectoire issue de $\mathbf{X}(0)$, pour $(R, P, bio)$. Nous pouvons voir tout cycle limite comme une trajectoire. Ainsi, tout cycle limite pour $(R, P, trad)$ est une suite-extraite d'un cycle limite pour $(R, P, bio)$. Afin de faciliter l'écriture et la lecture des équations suivantes, tout cycle $C$ de longueur $l$ que nous notions jusqu'à présent $C = [\mathbf{X}^0, \cdots, \mathbf{X}^{l-1}]$ est désormais noté $C = [\mathbf{X}^0, \cdots, \mathbf{X}^{l-1}, \mathbf{X}^l = \mathbf{X}^0]$.

Si $C_P^{trad} = [\mathbf{Y}^0, \cdots, \mathbf{Y}^{l-1}, \mathbf{Y}^l = \mathbf{Y}^0]$ est le cycle de longueur $l$ pour l'itération traditionnelle, et $C_P^{bio}$ le cycle correspondant pour l'itération bio-inspirée, alors, on peut écrire $C_P^{bio} = [\mathbf{X}^0, \cdots, \mathbf{X}^{k-1}, \mathbf{X}^k = \mathbf{X}^0]$ avec $\forall i \in \{0, \cdots, l\}, \mathbf{Y}^i = \mathbf{X}^{pi}$. On a donc $k = pl$. Comme $C_P^{bio}$ définit également les $k+1$ premières configurations de la trajectoire issue de $\mathbf{X}^0$ pour $(R, P, bio)$, on a $\forall j \in \{0, \cdots, k-1\},\ \mathbf{X}^{j+1} = G_F(P_{i_j}, \mathbf{X}^j)$, avec $i_j \in \{1, \cdots, p\},\ i_j \equiv j + 1\ [p]$. Nous pouvons donc définir les cycles limites pour l'itération bio-inspirée de la manière suivante.

**Définition 4.2** (Cycle limite pour l'itération bio-inspirée)
*Soit $R = (Q, F)$ un réseau d'automates à $n$ éléments, et $P = (P_1, \cdots, P_p)$ une partition ordonnée de l'ensemble $\{1, \cdots, n\}$.*
*La suite de configurations $C \subseteq Q^n, C = [\mathbf{X}^0, \cdots, \mathbf{X}^{l-1}, \mathbf{X}^l = \mathbf{X}^0]$ est un* cycle limite *de longueur $l$ pour $(R, P, bio)$, si et seulement, si $C$ vérifie :*

- $\exists k \in \mathbb{N}, k > 1\ /\ l = kp$,
- $\forall i, j \in \{1, \cdots, k\}, i \neq j,\ \mathbf{X}^{ip} \neq \mathbf{X}^{jp}$,
- $\forall j \in \{0, \cdots, l-1\},\ \mathbf{X}^{j+1} = G_F(P_{i_j}, \mathbf{X}^j)$, *avec* $i_j \in \{1, \cdots, p\},\ i_j \equiv j + 1\ [p]$.

Pour être tout à fait exact, dans la définition ci-dessus, il faut organiser l'écriture du cycle limite pour être certain d'avoir $\mathbf{X}^1 = G_F(P_1, \mathbf{X}^0)$. On sait que tout cycle limite $C = [\mathbf{X}^0, \cdots, \mathbf{X}^{l-1}, \mathbf{X}^l = \mathbf{X}^0]$ peut s'écrire $C = [\mathbf{X}^1, \cdots \mathbf{X}^{l-1}, \mathbf{X}^0, \mathbf{X}^1]$, ou $C = [\mathbf{X}^2, \cdots \mathbf{X}^{l-1}, \mathbf{X}^0, \mathbf{X}^1, \mathbf{X}^2]$... On suppose donc que l'on écrit toujours les cycles limites sous la forme $C = [\mathbf{X}^0, \cdots, \mathbf{X}^{l-1}, \mathbf{X}^0]$ avec $\mathbf{X}^1 = G_F(P_1, \mathbf{X}^0)$.

Dans l'exemple 4.1, la suite de configurations $C^{bio} = [000,\ 001,\ 011,\ 011,\ 010,$

000, 000] est bien un cycle limite de longueur 6 pour $(R, P, bio)$ avec $P = (3)(2)(1)$. $C^{bio}$ correspond au cycle limite $C^{trad} = [000, , 011, 000]$ de longueur 2, obtenu pour $(R, P, trad)$. On voit que $C^{trad}$ est une séquence extraite de $C^{bio}$. On dit donc que tout cycle limite obtenu pour l'itération traditionnelle est un cycle extrait du cycle limite correspondant pour l'itération bio-inspirée.

La définition des points fixes pour l'itération bio-simulée est la même que dans le cadre de l'itération traditionnelle. Une configuration $\mathbf{X} \in Q^n$ est un point fixe pour le réseau d'automates $R = (Q, F)$ si, et seulement si $F(\mathbf{X}) = \mathbf{X}$. Ainsi, le point fixe $\mathbf{X}$ vérifie $\forall E \in \mathcal{P}(\{1, \cdots, n\}), G_F(E, \mathbf{X}) = \mathbf{X}$. En particulier, quelle que soit la partition ordonnée $P = (P_1, \cdots, P_p)$ de $\{1, \cdots, n\}$, on a $\forall k \in \{1, \cdots, p\}$, $G_F(P_k, \mathbf{X}) = \mathbf{X}$. Les points fixes d'un réseau d'automates $R$ pour l'itération traditionnelle, et pour l'itération bio-inspirée, sont les mêmes. Soit $(\mathbf{X}^k)_{k\in\mathbb{N}}$ une trajectoire pour $(R, P, bio)$. Nous avons vu que deux configurations successives $\mathbf{X}^i$ et $\mathbf{X}^{i+1}$ de cette trajectoire peuvent être égales sans correspondre à un point fixe, $\mathbf{X}^i = \mathbf{X}^{i+1} = \mathbf{X}$. D'après la définition des points fixes ci-dessus, la configuration $\mathbf{X}$ ne peut se répéter plus de $p$ fois successivement dans la trajectoire $(\mathbf{X}^k)_{k\in\mathbb{N}}$.

Nous avons donc une correspondance totale entre les attracteurs d'un réseau d'automates pour l'itération traditionnelle, et pour l'itération bio-inspirée : chaque cycle limite pour un mode d'itération traditionnel est associé à un cycle limite pour le MIBSBI associé à la même partition ordonnée, et inversement ; les points fixes de l'itération traditionnelle et de l'itération bio-inspirée sont identiques. Ainsi, les classes de comportements que nous avons définies pour les RABS dans la partie 1.3.2 ne dépendent pas du choix entre itération traditionnelle et itération bio-inspirée.

Nous avons introduit l'itération bio-inspirée afin de mieux modéliser les réseaux de régulation génétique. Nous avons vu que, dans les trajectoires obtenues par itération bio-inspirée, deux configurations successives pouvaient être identiques sans correspondre à un point fixe. Si l'on considère que l'on peut observer, au moyen de bio-puces, l'évolution de l'expression des gènes d'un réseau correspondant à l'ensemble des configurations formant une trajectoire, il est impossible de distinguer deux configurations successives identiques. Ainsi le début de trajectoire $000 \to 001 \to 011 \to 011 \to 010 \to 000 \to 000$ sera observé comme étant identique à $000 \to 001 \to 011 \to 010 \to 000$. Afin de modéliser ce phénomène, nous appliquons une phase de réduction à toute trajectoire obtenue pour l'itération bio-inspirée.

Cette phase de réduction est définie comme suit. Soit $R$ un réseau d'automates à $n$ éléments, et $P$ une partition ordonnée de $\{1, \cdots, n\}$. A chaque configuration $\mathbf{X}^i$ d'une trajectoire $(\mathbf{X}^k)_{k\in\mathbb{N}}$ pour $(R, P, bio)$, on peut associer l'ensemble $E(\mathbf{X}^i) = \{j \in \mathbb{N}, j > i \ / \ \mathbf{X}^j \neq \mathbf{X}^i\}$. Par définition des points fixes, $E(\mathbf{X}^i) = \emptyset$ si, et seulement si $\mathbf{X}^i$ est un point fixe de $R$. On peut alors extraire de la trajectoire $(\mathbf{X}^k)_{k\in\mathbb{N}}$ la sous-suite $(\mathbf{X}^{\phi(k)})_{k\in\mathbb{N}}$ avec $\phi$ construite sur $\mathbb{N}$ par

$$\phi(k) = \begin{cases} 0 & \text{si } k = 0 \\ \min(E(\mathbf{X}^{\phi(k-1)})) & \text{si } k \neq 0 \text{ et } E(\mathbf{X}^{\phi(k-1)}) \neq \emptyset \\ k & \text{si } k \neq 0 \text{ et } E(\mathbf{X}^{\phi(k-1)}) = \emptyset \end{cases} .$$

Par construction, la fonction $\phi$ est croissante, donc la suite $(\mathbf{X}^{\phi(k)})_{k\in\mathbb{N}}$ est bien une suite extraite de $(\mathbf{X}^k)_{k\in\mathbb{N}}$. La suite de configurations $(\mathbf{X}^{\phi(k)})_{k\in\mathbb{N}}$ représente la trajectoire $(\mathbf{X}^k)_{k\in\mathbb{N}}$ pour $(R, P, bio)$, que l'on a réduite. Dans cette trajectoire réduite, deux configurations successives égales correspondent à un point fixe du réseau. La procédure de réduction affecte également les cycles limites. Toute séquence $[\mathbf{X}^0, \cdots, \mathbf{X}^{k-1}, \mathbf{X}^k = \mathbf{X}^0]$ obtenue par réduction du cycle limite $[\mathbf{X}^0, \cdots, \mathbf{X}^{l-1}, \mathbf{X}^l = \mathbf{X}^0]$ est appelée cycle limite réduit (de longueur $k$). En reprenant le cycle limite de l'exemple 4.1, le cycle limite réduit obtenu par réduction du cycle limite $C^{bio}$ = [000, 001, 011, 011, 010, 000, 000] est $C^{red}$ = [000, 001, 011, 010, 000].

Soit $R$ un réseau d'automates à $n$ éléments, $\mathbf{X}(0)$ une configuration de ce réseau, et $P$ une partition ordonnée de $\{1, \cdots, n\}$. Dans la trajectoire $(\mathbf{Y}^k)_{k\in\mathbb{N}}$ issue de $\mathbf{X}(0)$ (on a donc $\mathbf{Y}^0 = \mathbf{X}(0)$) pour $(R, P, trad)$, deux configurations successives $\mathbf{Y}^i$ et $\mathbf{Y}^{i+1}$ sont égales si, et seulement si $\mathbf{Y}^i$ est un point fixe du réseau $R$. Soit $(\mathbf{X}^k)_{k\in\mathbb{N}}$ la trajectoire issue de $X(0)$ pour $(R, P, bio)$. Nous avons vu que $(\mathbf{Y}^k)_{k\in\mathbb{N}}$ est une suite extraite de $(\mathbf{X}^k)_{k\in\mathbb{N}}$. Si l'on note $(\tilde{\mathbf{X}}^k)_{k\in\mathbb{N}}$ la trajectoire réduite obtenue par réduction de $(\mathbf{X}^k)_{k\in\mathbb{N}}$, on peut noter que, par construction de $(\tilde{\mathbf{X}}^k)_{k\in\mathbb{N}}$, $(\mathbf{Y}^k)_{k\in\mathbb{N}}$ est également une suite extraite de $(\tilde{\mathbf{X}}^k)_{k\in\mathbb{N}}$. De même tout cycle limite pour $(R, P, trad)$ est un cycle extrait d'un cycle limite réduit pour $(R, P, bio)$. En reprenant l'exemple 4.1, le cycle limite $C^{trad}$ = [000, , 011, 000] est bien un cycle extrait du cycle limite réduit $C^{red}$ = [000, 001, 011, 010, 000].

Nous pouvons formuler trois remarques concernant les cycles limites réduits :

- dans le cas où $R$ est itéré avec le mode parallèle, les itérations traditionnelle et bio-inspirée de $R$ sont identiques et toute trajectoire de $R$ pour l'itération bio-inspirée est invariante par le processus de réduction ; ainsi, pour le mode d'itération parallèle, les cycles limites en itération traditionnelle sont identiques aux cycles limites réduits en itération bio-inspirée ;
- dans le cas particulier où le cycle limite réduit pour $(R, P, bio)$, noté $C^{bio}$, est de longueur 2, le cycle limite extrait $C^{trad}$, qui correspond à l'itération traditionnelle $(R, P, trad)$, est identique à $C^{bio}$ (il ne peut pas être de longueur inférieure à 2).
- si $C = [\mathbf{X}^0, \cdots, \mathbf{X}^{l-1}, \mathbf{X}^l = \mathbf{X}^0]$ est un cycle limite réduit pour $R$ itéré avec un mode séquentiel bio-inspiré, alors on a $\forall k \in \{0, \cdots, l-1\}$, $d(\mathbf{X}^k, \mathbf{X}^{k+1}) = 1$, où $d$ désigne la distance de Hamming. En effet, pour un mode séquentiel bio-inspiré, à chaque pas de temps, l'état d'un seul élément du réseau est mis à jour. Ainsi, deux configurations successives d'une trajectoire réduite ne diffèrent que d' un seul élément.

Si $R$ est un réseau d'automates qui modélise un réseau de régulation génétique, et que $m$ est le mode d'itération blocs-séquentiel qui permet de modéliser les différents temps d'activation des gènes du réseau, le plus fidèlement possible, alors, les trajectoires réduites et cycles limites réduits de $R$ pour le mode $m$ en

itération bio-inspirée représentent l'évolution de l'expression des gènes du réseau, telle qu'on peut l'observer avec une bio-puce. Nous avons donc introduit et défini un processus d'itération des réseaux d'automates qui permet une meilleure modélisation des réseaux de régulation génétique.

Dans la suite, nous travaillons sur les cycles limites réduits en itération bio-inspirée. Nous démontrons certains résultats sur ces cycles réduits qui nous permettent de donner des propriétés relatives aux cycles limites en itération traditionnelle.

## 4.2 Résultats

Nous présentons tout d'abord une propriété des cycles limites de l'itération bio-inspirée qui montre que cette itération est plus adaptée à la modélisation des réseaux de régulation génétique que l'itération traditionnelle.

**Proposition 4.1**
*Soit $R = (Q, F)$ un réseau d'automates à $n$ éléments.*
*Soit $P = (P_1, \cdots, P_p)$ une partition ordonnée de $\{1, \cdots, n\}$.*
*Soit $P' = (P'_1, \cdots, P'_p)$ une partition ordonnée de $\{1, \cdots, n\}$ obtenue à partir de $P$ par décalage circulaire des blocs, c'est-à-dire $\exists i \in \{1, \cdots, p-1\}$, $/ \forall j \in \{1, \cdots, p\}$, $P'_j = P_k$, avec $k \in \{1, \cdots, p\}, k \equiv i + j \ [p]$.*

*Si $C \subseteq Q^n$ est un cycle limite pour $(R, P, bio)$, alors $C$ est un cyle limite pour $(R, P', bio)$.*

*Preuve.* On conserve les notations de la proposition.
On note $P' = (P'_1, \cdots, P'_p) = (P_i, \cdots, P_p, P_1, \cdots, P_{i-1})$.
On note $C = [\mathbf{X}^0, \cdots, \mathbf{X}^{l-1}, \mathbf{X}^l = \mathbf{X}^0]$, avec $\mathbf{X}^1 = G_F(P_1, \mathbf{X}^0)$.
Afin de ne pas alourdir la démonstration, on considère que les indices utilisés pour désigner les blocs des partitions ordonnées $P$ et $P'$ sont donnés modulo $p$ : $P_j = P_k$, avec $k \in \{1, \cdots, p\}, k \equiv j \ [p]$. On peut donc écrire, $\forall j \in \{1, \cdots, p\}$, $P'_j = P_{j+i-1}$. De même, les indices utilisés pour désigner les configurations du cycle limite $C$ sont donnés modulo $l$ : $X^j = X^k$, avec $k \in \{0, \cdots, l-1\}, k \equiv j \ [l]$.
$C$ étant un cycle limite pour $(R, P, bio)$, on peut donc écrire, $\forall j \in \{0, \cdots, l-1\}$, $\mathbf{X}^{j+1} = G_F(P_{j+1}, \mathbf{X}^j)$.
On s'intéresse à la trajectoire $(\mathbf{Y}^k)_{k \in \mathbb{N}}$, issue de $\mathbf{X}^{i-1}$, pour $(R, P', bio)$.
On montre par récurrence que $\forall k \in \mathbb{N}, \mathbf{Y}^k = \mathbf{X}^{k+i-1}$.
$(\mathbf{Y}^k)_{k \in \mathbb{N}}$ étant une trajectoire issue de $\mathbf{X}^{i-1}$, on a $\mathbf{Y}^0 = \mathbf{X}^{i-1}$.
On suppose que l'on a $(\mathbf{Y}^k) = \mathbf{X}^{k+i-1}$.
$(\mathbf{Y}^k)_{k \in \mathbb{N}}$ étant une trajectoire pour $(R, P', bio)$, on a $\mathbf{Y}^{k+1} = G_F(P'_{k+1}, \mathbf{Y}^k) = G_F(P_{k+i}, \mathbf{X}^{k+i-1}) = \mathbf{X}^{k+i}$.
On a donc bien, $\forall k \in \mathbb{N}, \mathbf{Y}^k = \mathbf{X}^{k+i-1}$.
$C$ étant un cycle limite, $C' = [\mathbf{X}^{i-1}, \cdots, \mathbf{X}^{l-1}, \cdots, \mathbf{X}^0, \cdots, \mathbf{X}^{i-1}]$ est un cycle limite pour $(R, P', bio)$, avec $\mathbf{X}^i = G_F(P'_1, \mathbf{X}^{i-1})$. On voit que l'on peut réécrire $C'$ sous la forme $C' = [\mathbf{X}^0, \cdots, \mathbf{X}^{l-1}, \mathbf{X}^0] = C$. Donc $C$ est bien un cycle limite

pour $(R, P', bio)$. □

Notons que cette propriété est valable pour les cycles limites non réduits de l'itération bio-inspirée. Mais si elle est valable pour un cycle limite non réduit, alors, elle est également valable pour le cycle limite réduit correspondant.

Dans l'itération bio-inspirée, une fois l'ordre des blocs établi, le choix du premier bloc n'a aucune influence sur les attracteurs du réseau. Dans le cadre de la modélisation de réseaux de régulation génétique, cette propriété est très importante car, comment déterminer, à partir d'un niveau d'expression génétique initial quel groupe de gènes va être le premier exprimé ? Notons que la propriété 4.1 n'est pas valable pour les cycles limites en itération traditionnelle.

*Exemple* 4.2 *:* Nous avons déjà vu que le réseau $R$ de l'exemple 4.1, itéré avec le mode d'itération séquentiel (3)(2)(1) avait un unique cycle limite de longueur 2 en itération traditionnelle : [000, 011, 000]. Toujours en itération traditionnelle, en itérant ce réseau avec le mode séquentiel (1)(3)(2), on obtient le même unique cycle limite. En revanche, avec le mode séquentiel (2)(1)(3), le réseau a également un unique cycle limite de longueur 2, mais différent : [001, 010, 001].
En itération bio-inspirée, pour les modes d'itération séquentiels (3)(2)(1), (2)(1)(3) et (1)(3)(2), le réseau a le même cycle limite de longueur 6 : [000, 001, 011, 011, 010, 000, 000], que l'on peut réduire à [000, 001, 011, 010, 000].

Nous rappelons que tout cycle limite pour $(R, P, trad)$ est un cycle extrait d'un cycle limite pour $(R, P, bio)$. La propriété 4.1 nous permet donc d'énoncer le corrollaire suivant, concernant les cycles limites de l'itération traditionnelle.

**Corollaire 4.1**
*Soit $R = (Q, F)$ un réseau d'automates à $n$ éléments.*
*Soit $P = (P_1, \cdots, P_p)$ une partition ordonnée de $\{1, \cdots, n\}$.*
*Soit $P' = (P'_1, \cdots, P'_p)$ une partition ordonnée de $\{1, \cdots, n\}$ obtenue à partir de $P$ par décalage circulaire des blocs.*

*Le nombre de cycles limites pour $(R, P, trad)$ est égal au nombre de cycles limites pour $(R, P', trad)$.*

Si l'on suppose qu'il y a un unique cycle limite pour $(R, P, trad)$ (et donc pour $(R, P', trad)$), alors les deux cycles limites correspondant à $(R, P, trad)$ et $(R, P', trad)$ sont deux cycles extraits d'un même cycle. Ils peuvent être différents lorsqu'ils sont extraits de façon déphasée, comme le montre la figure 4.2. Cette figure correspond aux cycles limites présentés dans l'exemple 4.2. A partir du cycle limite réduit pour $R$ itéré avec le mode d'itération séquentiel (3)(2)(1) pour l'itération bio-inspirée, on trouve les cycles limites extraits qui correspondent aux modes d'itération séquentiels (3)(2)(1), (2)(1)(3) et (1)(3)(2) pour l'itération traditionnelle.

Avant d'énoncer et de prouver le résultat suivant, il nous faut introduire quelques définitions supplémentaires, nous permettant de faire le lien entre les partitions de $\{1, \cdots, n\}$ et les cycles limites.

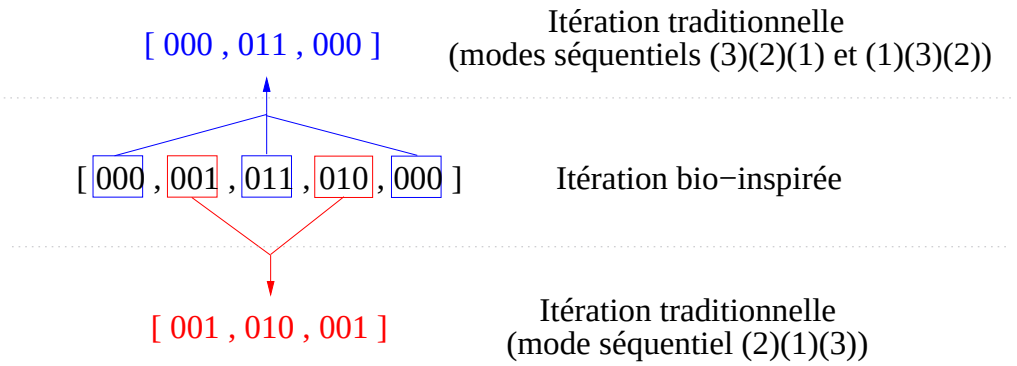



FIG. 4.2 – Schéma représentant l'extraction des cycles limites pour l'itération traditionnelle à partir d'un cycle limite réduit pour l'itération bio-inspirée.

Dans la partie 3.2, nous avons défini une relation d'ordre sur les partitions ordonnées. Nous allons ici définir une relation d'ordre sur les partitions non ordonnées.

**Définition 4.3** (Relation d'ordre sur les partitions non ordonnées)
*Soient* $P^1 = (P^1_1, \cdots, P^1_{p_1})$ *et* $P^2 = (P^2_1, \cdots, P^2_{p_2})$ *deux partitions non ordonnées de l'ensemble* $\{1, \cdots, n\}$. *On dit que* $P_2$ *est* plus fine *que* $P_1$ *si, et seulement si, on a pour tout couple* $(i, j) \in \{1, \cdots, n\}^2$ *:*

$$\exists k \in \{1, \cdots, p_1\} \,/\, i \in P^1_k, j \notin P^1_k \;\Rightarrow\; \exists l \in \{1, \cdots, p_2\} / i \in P^2_l, j \notin P^2_l$$

La relation "plus fine" étant définie pour les partitions ordonnées, nous pouvons définir la partition liée à un cycle limite, que nous appelons partition cyclique.

**Définition 4.4** (Partition cyclique)
*Soit* $C = [\mathbf{X}^0, \cdots, \mathbf{X}^{l-1}, \mathbf{X}^l = \mathbf{X}^0] \subseteq Q^n$ *un cycle de longueur* $l$.
*On appelle* partition cyclique *associée au cycle* $C$, *la partition non ordonnée de l'ensemble* $\{1, \cdots, n\}$, *notée* $P = (P_1, \cdots, P_p)$, *qui est la partition la plus fine telle que, pour tout couple* $(i, j) \in \{1, \cdots, n\}^2$ *:*

$$\exists k \in \{1, \cdots, l\} \,/\, \left\{ \begin{array}{l} x^{k-1}_i \neq x^k_i \\ x^{k-1}_j \neq x^k_j \end{array} \right. \Rightarrow\; \exists m \in \{1, \cdots, p\} \,/\, \left\{ \begin{array}{l} i \in P_m \\ j \in P_m \end{array} \right.$$

De façon informelle, si $R$ est un réseau d'automates et $C$ un cycle limite atteint par $R$, la partition cyclique associée au cycle $C$ est la partition non ordonnée la plus fine dont les blocs regroupent les éléments de $R$ qui changent d'état au même moment dans $C$.

On voit que la partition cyclique associée à un cycle limite $C$ pour un MIBSBI donné est identique à la partition cyclique au cycle limite réduit correspondant à $C$. En effet, la définition de la partition cyclique se fait à partir des transitions

entre deux configurations distinctes. Ces transitions sont les mêmes entre un cycle limite pour MIBSBI donné et le cycle limite réduit correspondant.

L'exemple suivant montre comment construire la partition cyclique associée à un cycle donné.

*Exemple* 4.3 *:* On considère le réseau d'automates booléens à 3 éléments, $R$, présenté dans l'exemple 4.1. Ce réseau, itéré avec le mode parallèle (nous avons vu que dans ce cas, les itérations traditionnelle et bio-inspirée sont identiques), atteint le cycle limite de longueur 4 suivant : $C = [001, 011, 010, 000, 001]$. On constate qu'entre 2 configurations successives de $C$, l'état d'un seul élément change. Ainsi, la partition cyclique associée à $C$ est la partition $P = (\{1\}, \{2\}, \{3\})$, que l'on note aussi plus simplement $P = (1)(2)(3)$.
On considère maintenant le réseau d'automates booléens à 3 éléments qui modélise la morphogénèse de la plume chez le poulet, $R_{plume}$, présenté dans l'exemple 1.5 de la partie 1.3.2. Nous avons vu que ce réseau, itéré avec le mode parallèle, atteint le cycle de longueur 4 suivant : $C_{plume} = [000, 100, 111, 010, 000]$. Entre la deuxième configuration de $C_{plume}$, 100, et la troisième, 111, l'état des éléments 2 et 3 du réseau change. 2 et 3 sont donc dans un même bloc de la partition cyclique associée à $C_{plume}$. Entre 111 et 010, l'état des éléments 1 et 3 est modifié. 1 et 3 sont donc dans un même bloc de la partition cyclique associée à $C_{plume}$. On en déduit que la partition cyclique associée à $C_{plume}$ est la partition $P_{plume} = (1, 2, 3)$.

**Lemme 4.1**
*Soit $R = (Q, F)$ un réseau d'automates à $n$ éléments.*
*Soit $P = (P_1, \cdots, P_p)$ une partition ordonnée de $\{1, \cdots, n\}$.*
*Soit $C = [\mathbf{X}^0, \cdots, \mathbf{X}^{l-1}, \mathbf{X}^l = \mathbf{X}^0] \subseteq Q^n$ un cycle.*
*Soient les 2 propositions suivantes :*

1. *$C$ est un cycle limite pour $(R, P, bio)$ ;*
2. *la partition cyclique associée à $C$ est plus fine que la partition non ordonnée construite avec les mêmes blocs que $P$.*

*On a : 1 $\Rightarrow$ 2.*

*Preuve.* On suppose que la partition cyclique associée à $C$ n'est pas plus fine que la partition non ordonnée construite avec les mêmes blocs que $P$.
Cela signifie que $P$ est telle qu'il existe 2 éléments $i_1$ et $i_2$ qui changent d'état au même moment dans $C$ (entre $\mathbf{X}^k$ et $\mathbf{X}^{k+1}$, avec $k \in \{0, \cdots, l-1\}$), tels que $i_1$ et $i_2$ ne sont pas dans un même bloc de $P$. On note $P_m$ le bloc de $P$ qui contient $i_1$.
Soit $\mathbf{Y} = G_F(P_m, \mathbf{X}^k)$. Comme $i_2 \notin P_m$, alors $y_{i_2} = x^k_{i_2}$, et donc $y_{i_2} \neq x^{k+1}_{i_2}$, soit $\mathbf{Y} \neq \mathbf{X}^{k+1}$.
Soit $j \in \{1, \cdots, p\}$, $j \neq m$, et $\mathbf{Y} = G_F(P_j, \mathbf{X}^k)$. Comme $i_1 \notin P_j$, alors $y_{i_1} = x^k_{i_1}$, et donc $y_{i_1} \neq x^{k+1}_{i_1}$, soit $\mathbf{Y} \neq \mathbf{X}^{k+1}$.
Ainsi, $\forall j \in \{1, \cdots, p\}, G_F(P_j, \mathbf{X}^k) \neq \mathbf{X}^{k+1}$. La transition entre $\mathbf{X}^k$ et $\mathbf{X}^{k+1}$ ne peut donc pas se faire pour $(R, P, bio)$. $C$ n'est donc pas un cycle limite pour $(R, P, bio)$. □

A partir du lemme 4.1, on peut donc définir les partitions compatibles avec un cycle.

**Définition 4.5** (Partition compatible avec un cycle)
*Soit* $P = (P_1, \cdots, P_p)$ *une partition ordonnée de* $\{1, \cdots, n\}$, *et* $C = [\mathbf{X}^0, \cdots, \mathbf{X}^{l-1}, \mathbf{X}^l = \mathbf{X}^0] \subseteq Q^n$ *un cycle.*
*On dit que* $P$ *est une* partition compatible *avec le cycle* $C$ *si, et seulement si la partition cyclique associée à* $C$ *est plus fine que la partition non ordonnée construite avec les mêmes blocs que* $P$.

*Exemple* 4.4 *:* Nous avons vu dans l'exemple précédent que la partition cyclique associée au cycle $C = [001, 011, 010, 000, 001]$ est la partition $P = (1)(2)(3)$. Ainsi toute partition ordonnée de $\{1, 2, 3\}$ est une partition compatible avec le cycle $C$.

**Théorème 4.1**
*Soit* $R = (Q, F)$ *un réseau d'automates à* $n$ *éléments.*
*Soit* $P = (P_1, \cdots, P_p)$ *une partition ordonnée de* $\{1, \cdots, n\}$.
*Soit* $C = [\mathbf{X}^0, \cdots, \mathbf{X}^{l-1}, \mathbf{X}^l = \mathbf{X}^0] \subseteq Q^n$ *un cycle limite pour* $R$ *itéré avec le mode parallèle.*

*Les 2 propositions suivantes sont équivalentes :*
- $C$ *est un cycle limite réduit pour* $(R, P, bio)$ *;*
- $P$ *est compatible avec* $C$

*Preuve.* Le lemme 4.1 fournit un sens de l'équivalence. Il faut donc montrer que si $P$ est compatible avec $C$, alors $C$ est un cycle limite réduit pour $(R, P, bio)$.
$C$ est un cycle limite pour le mode parallèle, donc, pour tout $k \in \{0, \cdots, l-1\}$, $F(\mathbf{X}^k) = \mathbf{X}^{k+1}$.
On définit l'ensemble $I^k = \{i \in \{1, \cdots, n\} \quad ; \quad x_i^k \neq x_i^{k+1}\}$, pour tout $k \in \{0, \cdots, l-1\}$.
Comme $P$ est compatible avec $C$, alors pour tout $k \in \{0, \cdots, l-1\}$ il existe un unique $j_k \in \{1, \cdots, p\}$ tel que $I^k \subseteq P_{j_k}$, et pour tout $j \in \{1, \cdots, p\}, j \neq j_k$, on a $I^k \cap P_j = \emptyset$. On a alors, pour tout $k \in \{0, \cdots, l-1\}$,
$F_G(P_j, \mathbf{X}^k) = \begin{cases} \mathbf{X}^{k+1} & \text{si } j = j_k, \\ \mathbf{X}^k & \text{sinon} \end{cases}$.
On en déduit donc que $C$ est un cycle limite réduit pour $(R, P, bio)$. □

Ce résultat est important : il établit un lien entre les cycles limites du mode d'itération parallèle et les cycles limites des autres modes d'itération blocs-séquentiels, pour l'itération bio-inspirée. Il permet de déterminer, en fonction du nombre d'éléments dont la valeur change en même temps dans un cycle limite pour le réseau itéré avec le mode parallèle, pour quels MIBSBI le cycle est également un cycle limite réduit.

Si une partition ordonnée $P$, n'est compatible avec aucun cycle limite pour le réseau $R$ itéré avec le mode parallèle, il peut tout de même exister un cycle limite pour $(R, P, bio)$, comme le montre l'exemple suivant.

*Exemple* 4.5 *:* On considère le réseau $R' = (\mathbf{W}', \boldsymbol{\theta}')$ dont la matrice d'interaction et le vecteur seuil sont donnés ci-après :

$$\mathbf{W}' = \begin{pmatrix} -1 & -1 & 0 \\ -1 & 0 & -1 \\ 0 & -1 & -1 \end{pmatrix} \qquad \boldsymbol{\theta}' = \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix}$$

Itéré avec le mode parallèle, $R'$ a deux cycles limites de longueur 2 : $C_1 =$ [000, 111, 000] et $C_2 =$ [001, 100, 001]. Les partitions cycliques associées à $C_1$ et $C_2$ sont respectivement $P_1 = (1, 2, 3)$ et $P_2 = (1, 3)(2)$. La partition $P = (1)(2)(3)$ n'est donc compatible ni avec $C_1$, ni avec $C_2$. Pourtant, on observe un cycle limite pour $(R', P, bio)$ dont le cycle limite réduit correspondant est : $C_3 =$ [000, 100, 101, 001, 000].

A partir du théorème 4.1, on obtient notamment le corollaire suivant.

**Corollaire 4.2**
*Soit $R = (Q, F)$ un réseau d'automates à $n$ éléments.*
*Soit $P = (P_1, \cdots, P_p)$ une partition ordonnée de $\{1, \cdots, n\}$.*
*Soit $C = [\mathbf{X}^0, \cdots, \mathbf{X}^{l-1}, \mathbf{X}^l = \mathbf{X}^0] \subseteq Q^n$ un cycle limite pour $R$ itéré avec le mode parallèle.*

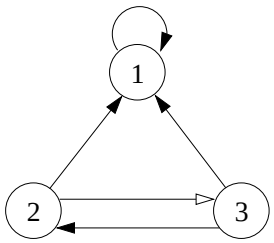
*Les 2 propositions suivantes sont équivalentes :*
- *$C$ est tel que $\forall k \in \{0, \cdots, l-1\}$, $d(\mathbf{X}^k, \mathbf{X}^{k+1}) = 1$ (d, distance de Hamming) ;*
- *$C$ est un cycle limite réduit pour $R$ itéré avec n'importe quel mode d'itération blocs-séquentiel bio-inspiré.*

*Preuve.* Si $C$ est tel que $\forall k \in \{0, \cdots, l-1\}$, $d(\mathbf{X}^k, \mathbf{X}^{k+1}) = 1$, alors la partition cyclique associée à $C$ est la partition non ordonnée $P = \{\{i\},\ i \in \{1, \cdots, n\}\}$. D'après le théorème 4.1, $C$ est donc un cycle limite réduit pour $R$ itéré avec n'importe quel MIBSBI.
Si $C$ est un cycle limite réduit pour $R$ itéré avec n'importe quel MIBSBI, alors, en particulier, $C$ est un cycle limite réduit pour $R$ itéré avec un mode séquentiel bio-inspriré. Or, nous avons remarqué à la fin de la partie 4.1, que dans ce cas, on a $\forall k \in \{0, \cdots, l-1\}$, $d(\mathbf{X}^k, \mathbf{X}^{k+1}) = 1$. □

D'après ce corollaire, on peut donc avoir, en itération bio-inspirée, un même cycle limite réduit pour tous les modes d'itération blocs-séquentiels. D'après ce corollaire, le théorème suivant, démontré par Goles et Salinas [79, 81], n'est pas valable en itération bio-inspirée, comme le montre l'exemple 4.6.

**Théorème 4.2** ([79, 81])
*Soit $R = (\{0, 1\}, F)$ un réseau d'automates booléens tel que les boucles du graphe d'interaction sont toutes positives.*
*Si $C = [\mathbf{X}^0, \cdots, \mathbf{X}^{l-1}, \mathbf{X}^l = \mathbf{X}^0] \subseteq \{0, 1\}^n$ est un cycle limite pour $R$ itéré avec le mode parallèle, alors $C$ n'est pas un cycle limite pour $R$ itéré avec un mode d'itération séquentiel.*

*Exemple* 4.6 *:* On reprend le réseau d'automates booléens à seuil à 3 éléments $R = (\mathbf{W}, \boldsymbol{\theta})$ de l'exemple 4.1, dont on rappelle la matrice d'interaction et le vecteur seuil, et dont on donne le graphe d'interaction, figure 4.3. On voit que la seule boucle de ce graphe est positive.




$$\mathbf{W} = \begin{pmatrix} 2 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \quad \boldsymbol{\theta} = \begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix}$$

FIG. 4.3 – Graphe d'interaction du réseau d'automates booléens à seuil des exemples 4.1 et 4.6.

$R$ est un réseau de la classe de comportement $Cy$ (nous avons vu précédemment que les classes de comportement ne dépendaient pas de l'itération choisie, traditionnelle ou bio-inspirée). Pour certains modes d'itération blocs-séquentiels, on observe un unique cycle limite, pour d'autres, on en observe 2 (le nombre de cycles limites ne dépend pas non plus de l'itération choisie). Le tableau 4.1 donne, pour chaque partition ordonnée $P$ de $\{1, 2, 3\}$, le(s) cycle(s) limite(s) pour $(R, P, trad)$, ainsi que le(s) cycle(s) limite(s) réduit(s) pour $(R, P, bio)$.

Le cycle limite observé pour $R$ itéré avec le mode parallèle, $[001, 011, 010, 000, 001]$, est tel que, à chaque transition, l'état d'un seul élément du réseau est modifié. On voit que $[001, 011, 010, 000, 001]$ est un cycle limite réduit pour $R$ itéré avec n'importe quel MIBSBI. En revanche, $[001, 011, 010, 000, 001]$, cycle limite pour le mode parallèle, n'est pas un cycle limite pour tous les modes d'itération blocs-séquentiels de l'itération traditionnelle, comme le prouve le théorème 4.2.

Nous avons vu l'importance du théorème 4.1 pour les cycles limites réduits de l'itération bio-inspirée. Ce théorème nous permet également d'énoncer certains résultats concernant les cycles limites de l'itération traditionnelle.

**Corollaire 4.3**
*Soit $R = (Q, F)$ un réseau d'automates à $n$ éléments.*

*Pour tout cycle limite $C = [\mathbf{X}^0, \cdots, \mathbf{X}^{l-1}, \mathbf{X}^l = \mathbf{X}^0] \subseteq Q^n$ pour $R$ itéré avec le mode parallèle, il existe un cycle limite pour $(R, P, trad)$, avec $P$, partition ordonnée de $\{1, \cdots, n\}$ compatible avec $C$. Ce cycle limite est un cycle limite extrait de $C$.*

*Preuve.* Il suffit de voir que tout cycle pour $(R, P, trad)$ est un cycle extrait d'un cycle limite réduit pour $(R, P, bio)$ et d'appliquer le théorème 4.1. □

Ce corollaire nous permet de donner une caractérisation des éventuels cycles limites pour un réseau de la classe $Ev$, itéré avec le mode parallèle : si $C =$

TAB. 4.1 – Cycles limites pour l'itération traditionnelle et cycles limites réduits pour l'itération bio-inspirée, observés pour le réseau d'automates booléens à seuil des exemples 4.1 et 4.6.

| Mode d'itération | Itération traditionnelle | Itération bio-inspirée |
|---|---|---|
| $(1,2,3)$ | $[001,011,010,000,001]$ | $[001,011,010,000,001]$ |
| $(1,2)(3)$ | $[001,010,001]$<br>$[110,101,110]$ | $[001,011,010,000,001]$<br>$[110,100,101,111,110]$ |
| $(1,3)(2)$ | $[011,000,011]$ | $[001,011,010,000,001]$ |
| $(2,3)(1)$ | $[001,011,010,000,001]$ | $[001,011,010,000,001]$ |
| $(1)(2,3)$ | $[001,011,010,000,001]$ | $[001,011,010,000,001]$ |
| $(2)(1,3)$ | $[001,010,001]$ | $[001,011,010,000,001]$ |
| $(3)(1,2)$ | $[011,000,011]$<br>$[100,111,100]$ | $[001,011,010,000,001]$<br>$[110,100,101,111,110]$ |
| $(1)(2)(3)$ | $[001,010,001]$<br>$[110,101,110]$ | $[001,011,010,000,001]$<br>$[110,100,101,111,110]$ |
| $(1)(3)(2)$ | $[011,000,011]$ | $[001,011,010,000,001]$ |
| $(2)(1)(3)$ | $[001,010,001]$ | $[001,011,010,000,001]$ |
| $(2)(3)(1)$ | $[001,010,001]$<br>$[110,101,110]$ | $[001,011,010,000,001]$<br>$[110,100,101,111,110]$ |
| $(3)(1)(2)$ | $[011,000,011]$<br>$[100,111,100]$ | $[001,011,010,000,001]$<br>$[110,100,101,111,110]$ |
| $(3)(2)(1)$ | $[011,000,011]$ | $[001,011,010,000,001]$ |

$[\mathbf{X}^0,\cdots,\mathbf{X}^{l-1},\mathbf{X}^l=\mathbf{X}^0]$ est cycle limite pour un réseau de la classe *Ev*, itéré avec le mode parallèle, alors, $\exists k\in\{0,\cdots,l-1\}\ /\ d(\mathbf{X}^k,\mathbf{X}^{k+1})>1$. En effet, dans le cas contraire, pour tout mode d'itération blocs-séquentiel traditionnel, il existe un cycle limite, et donc le réseau est un réseau de la classe *Mi* (on considère qu'il existe au moins un point fixe pour le réseau).

On peut enfin énoncer un dernier résultat relatif aux cycles limites pour l'itération traditionnelle, sous la forme du corollaire suivant.

**Corollaire 4.4**
*Soit $R=(Q,F)$ un réseau d'automates à $n$ éléments.*
*Soit $C=[\mathbf{X}^0,\cdots,\mathbf{X}^{l-1},\mathbf{X}^l=\mathbf{X}^0]\subseteq Q^n$ un cycle limite pour $R$ itéré avec le mode parallèle.*

*Soit $P$ une partition ordonnée de $\{1,\cdots,n\}$, compatible avec $C$.*

*Si $C$ est de longueur 2, alors, pour toute partition ordonnée de $\{1,\cdots,n\}$ compatible avec $C$, $C$ est un cycle limite pour $(R,P,trad)$.*

*Preuve.* Ce corollaire est une conséquence directe du corollaire 4.3, et du fait que si $C$ est un cycle limite réduit de longueur 2 pour $(R, P, bio)$, alors $C$ est un cycle limite de longueur 2 pour $(R, P, trad)$, comme nous l'avions remarqué à la fin de la partie 4.1. □

*Exemple* 4.7 *:* On considère $R_{Men}$, le réseau d'automates booléens à seuil qui modélise le régulation génétique lié à la morphogénèse florale chez *Arabidopsis thaliana*, présenté dans la partie 3.3. Nous avons vu que l'on peut déterminer toute la dynamique de ce réseau de 12 éléments en ré-écrivant la matrice d'interaction et le vecteur seuil donnés par Mendoza et Alvarez-Buylla avec des vecteurs d'interaction minimaux. Ceci nous permet de nous ramener à l'étude de 2 sous-réseaux de taille 3.
Nous allons, ici, considérer le réseau de taille 12 sans en modifier la matrice d'interaction et le vecteur seuil. Si l'on itère ce réseau avec le mode parallèle, on trouve les 6 points fixes présentés dans la partie 3.3, ainsi que 7 cycles limites que nous rappelons :

- $C_1 = [110000010000,\ 110000000110]$ ;
- $C_2 = [000100010000,\ 000100000110]$ ;
- $C_3 = [000000011000,\ 000000001110]$ ;
- $C_4 = [000100011110,\ 000000010110]$ ;
- $C_5 = [000100001000,\ 000000000000]$ ;
- $C_6 = [000100011000,\ 000000000110]$ ;
- $C_7 = [000100001110,\ 000000010000]$.

Tous ces cycles limites sont de longueur 2. Nous pouvons calculer la partition cyclique associée à chacun d'eux. On trouve alors 3 partitions cycliques différentes :

- $P_1 = (8, 10, 11)(1)(2)(3)(4)(5)(6)(7)(9)(12)$, associée aux cycles limites $C_1$, $C_2$ et $C_3$ ;
- $P_2 = (4, 9)(1)(2)(3)(5)(6)(7)(8)(10)(11)(12)$, associée aux cycles limites $C_4$ et $C_5$ ;
- $P_3 = (4, 8, 9, 10, 11)(1)(2)(3)(5)(6)(7)(12)$, associée aux cycles limites $C_6$ et $C_7$.

D'après le corollaire 4.4, pour l'itération traditionnelle, on peut donc affirmer que :

- les cycles limites $C_1$, $C_2$ et $C_3$ sont des cycles limites pour $R_{Men}$ itéré avec tout mode d'itération blocs-séquentiel tel que les éléments 8, 10 et 11 sont dans un même bloc ;
- les cycles limites $C_4$ et $C_5$ sont des cycles limites pour $R_{Men}$ itéré avec tout mode d'itération blocs-séquentiel tel que les éléments 4 et 9 sont dans un même bloc ;
- les cycles limites $C_6$ et $C_7$ sont des cycles limites pour $R_{Men}$ itéré avec tout mode d'itération blocs-séquentiel tel que les éléments 4, 8, 9, 10 et 11 sont dans un même bloc.

En remarquant que les éléments 4, 8, 9, 10 et 11 du réseau $R_{Men}$ sont respectivement AP1, BFU, AG, AP3 et PI, on retrouve la distribution des cycles limites en fonction des modes d'itération donnée dans la partie 3.3.

Ainsi, en tenant compte uniquement des cycles limites pour le réseau itéré avec le mode parallèle, et sans simuler la dynamique de ce réseau pour un autre mode d'itération, on retrouve les modes d'itération blocs-séquentiels de l'itération traditionnelle pour lesquels au moins l'un des 7 cycles limites est présent. Si l'on cherche les modes d'itérations blocs-séquentiels pour lesquels on n'a pas de cycle limite (rappelons que ces cycles limites n'ont aucune signification biologique pour ce réseau de régulation), les corollaires 4.3 et 4.4 nous permettent d'éliminer rapidement les mauvais candidats.

Dans ce chapitre, nous avons introduit une nouvelle itération, qui nous permet de modéliser, de façon plus fidèle que l'itération traditionnelle, la dynamique des réseaux de régulation génétique, à l'aide d'un réseau d'automates. Cette itération, que nous avons appelée itération bio-inspirée, est valable pour tout réseau d'automates, et non pas uniquement pour les réseaux d'automates booléens à seuil. Nous avons introduit les notions de trajectoires et cycles limites réduits pour un réseau, qui représentent l'évolution de l'expression génétique telle que l'on peut l'observer expérimentalement, au moyen de bio-puces.

A partir de la définition des cycles limites réduits, nous avons pu démontrer quelques résultats sur le lien existant entre ces cycles limites et les modes d'itération blocs-séquentiels bio-inspirés. Ces résultats nous permettent notamment de décrire, à partir des cycles limites observés pour un réseau itéré avec un mode parallèle, l'ensemble des modes d'itération blocs-séquentiels bio-inspirés pour lesquels ces cycles sont également des cycles limites pour le réseau.

# Conclusion et perspectives

Nous allons clore ce document en dressant un bilan de notre travail de doctorat. Nous allons présenter nos principales contributions, puis évoquer les problèmes ouverts qui nous semblent les plus intéressants, concernant la robustesse des réseaux de régulation aux modes d'itération.

## Conclusion

L'objet de notre travail était la compréhension de l'influence des modes d'itération sur la dynamique des réseaux de régulation. Plus précisément, nous avons étudié la conséquence de perturbations des modes d'itération sur les attracteurs d'un réseau de régulation, modélisé par un réseau d'automates booléens à seuil (dans la suite, nous utilisons la notation RABS, comme dans le corps du document). Nous avons abordé cette problématique de façon théorique, mais aussi au moyen de simulations numériques. Cette étude nous a également permis de nous interroger sur la validité de la modélisation de réseaux de régulation génétique sous la forme de RABS.

Dans le chapitre 1, après avoir introduit l'ensemble des notions utiles à la compréhension de notre étude, nous avons défini, dans la partie 1.3.2, 4 classes de comportement dynamique pour les réseaux d'automates (ces classes sont valables pour tous les réseaux d'automates, et pas uniquement pour les RABS). Cette classification est définie à partir de la robustesse des réseaux, au niveau de leurs attracteurs, face aux perturbations des modes d'itération blocs-séquentiels. Nous rappelons les 4 classes de comportement :

- les réseaux de la classe *Cy* sont les réseaux d'automates pour lesquels on n'observe que des cycles limites, quel que soit le mode d'itération blocs-séquentiels avec lequel ils sont itérés ;
- les réseaux de la classe *Fi* sont les réseaux pour lesquels on n'observe que des points fixes, quel que soit le mode d'itération ;
- les réseaux de la classe *Mi* sont les réseaux pour lesquels on observe au moins un point fixe et au moins un cycle limite, quel que soit le mode d'itération ;
- enfin, les réseaux de la classe *Ev* sont les réseaux pour lesquels on observe toujours au moins un point fixe, quel que soit le mode d'itération, mais pour lesquels, en fonction du mode d'itération, on observe, ou pas, au moins un

cycle limite.
Cette classification nous semble importante, car elle nous permet de distinguer les réseaux d'automates qui sont robustes aux perturbation des modes d'itération, de ceux qui, au contraire, ont un comportement dynamique qui est très sensible à ces perturbations. Nous avons pu étudier la robustesse des RABS, à partir de ces 4 classes de comportement dynamique, ce qui nous a permis de donner des résultats globaux, classe par classe, et non pas simplement, réseau par réseau.

Dans le chapitre 2, nous avons ensuite montré comment les RABS étaient construits à partir de l'ensemble des fonctions booléennes à seuil. Nous avons notamment insisté sur le fait qu'il fallait définir les réseaux à partir des matrices d'interaction et vecteurs seuils en utilisant des vecteurs d'interaction minimaux. Ces vecteurs garantissent que chaque poids d'interaction non nul d'un RABS représente effectivement une interaction entre les gènes du réseau de régulation génétique modélisé. Si le poids est négatif (resp. positif), l'interaction est de type inhibition (resp. activation). Dans ce chapitre, notre principale contribution concerne la partie 2.3, dans laquelle nous avons fait l'énumération des RABS à $n$ éléments, avec $1 \leq n \leq 7$. Nous avons défini une relation d'équivalence sur les RABS telle que, sont regroupés dans une même classe d'équivalence les réseaux qui sont identiques, à une permutation près de leurs éléments. Nous avons fait appel à la théorie des actions de groupe, et plus particulièrement à la formule de Burnside, pour dénombrer l'ensemble des classes d'équivalence de RABS. Nous avons également montré comment énumérer les représentants de toutes les classes d'équivalence de RABS. Cette énumération nous a permis de connaître la structure de la population totale de ces réseaux, et donc d'en construire des échantillons représentatifs, utiles pour les simulations numériques.

Le chapitre 3, qui présente l'ensemble des résultats obtenus par simulations numériques, est l'une de nos contribution majeures. Nous y avons tout d'abord montré que, pour les RABS de taille inférieure à 7, la robustesse générale diminuait avec la taille. Nous avons notamment montré que, lorsque la taille des réseaux augmentait, la proportion des réseaux de la classe de comportement *Fi* tendait vers 0, alors que la proportion des réseaux les moins robustes, ceux de la classe *Ev* augmentait jusqu'à atteindre 40% de la population totale des RABS à 7 éléments. Cette diminution de la robustesse avec l'augmentation de la taille des réseaux est un fait important dans le cadre de la modélisation des réseaux de régulation génétique. En effet, les réseaux de régulation biologique, de façon générale, se doivent d'être relativement robustes, afin de continuer à assurer leur fonction, malgré des perturbations de leur environnement, ou même de leur structure interne. Nous avons montré qu'en moyenne, les réseaux de régulation génétique pouvaient être plus robustes que les RABS de même taille, du fait de leur connectivité. En effet, les résultats de nos simulations numériques tendent à montrer que la baisse de la robustesse avec l'augmentation de la taille des RABS, est due, en partie, à une augmentation de la connectivité des réseaux (la forte connectivité, seule, ne permettant pas d'expliquer complètement la baisse

de la robustesse). La faible connectivité des réseaux de régulation génétique leur assurerait une plus grande robustesse aux perturbations des modes d'itération. Nous nous sommes particulièrement intéressés aux RABS de la classe de comportement *Ev*. Nous avons proposé d'expliquer l'apparition et la disparition des cycles limites par une (dé)synchronisation de certains éléments du réseau. Nous avons montré, toujours au moyen de simulations, que plus la taille des réseaux augmentait, moins cette explication était suffisante. Il semble que l'ordre de mise à jour des éléments joue également un rôle important dans l'apparition et la disparition de ces cycles limites. Enfin, toujours dans ce chapitre 3, nous avons abordé deux sujets plus directement liés à la modélisation des réseaux de régulations génétique. Nous avons tout d'abord traité l'exemple du réseau de régulation génétique lié à la morphogénèse florale chez *Arabidopsis thaliana*. Cet exemple nous a permis de montrer toute l'importance de l'utilisation de vecteurs d'interaction minimaux pour construire les matrices d'interaction et vecteurs seuils des RABS modélisant des réseaux de régulation. Nous avons également montré que, grâce à une bonne compréhension du modèle, on pouvait étudier la dynamique complète de ce réseau comportant 12 éléments, en se ramenant à l'étude de 2 sous-réseaux de taille 3. Cette simplification démontre, d'une part, qu'il est important de bien comprendre le modèle afin de pouvoir tirer des conclusions sur le réseau de régulation modélisé, mais aussi, justifie notre étude exhaustive des RABS de petite taille, qui représentent des modules des réseaux de grande taille. Nous avons, enfin, abordé le rôle des $miARN$ dans la régulation génétique. Au moyen de simples simulations, nous avons montré que ces $miARN$ permettaient de rendre les réseaux de régulation plus robustes aux perturbations des modes d'itération. Ce résultat est un premier résultat prometteur, car il permet d'entrevoir quels peuvent être les éléments du processus de régulation qui assurent la robustesse des réseaux de régulation génétique.

Le chapitre 4 concerne les réseaux d'automates de façon générale, et pas uniquement les RABS. Deux points particuliers de ce dernier chapitre nous semblent importants. Nous avons tout d'abord proposé de modifier l'itération des réseaux d'automates, afin de modéliser au mieux la régulation génétique. Nous avons introduit une itération, que nous avons appelée itération bio-inspirée. Dans cette itération, les trajectoires et cycles limites, et plus précisément, les trajectoires réduites et cycles limites réduits modélisent l'évolution, observable au moyen de bio-puces, de l'expression des gènes d'un réseau de régulation. Dans le processus de mise au point et de perfectionnement d'un modèle, l'étape de remise en cause du modèle est obligatoire. L'introduction de l'itération bio-inspirée est, à notre avis, un point important dans cette remise en cause du modèle. Nous considérons que la modélisation des réseaux de régulation génétique sous forme de réseaux d'automates itérés avec l'itération bio-inspirée est plus appropriée qu'avec l'itération traditionnelle. Nous insistons sur le fait que les 4 classes de comportement dynamique, que nous avons définies dans la partie 1.3.2, ne dépendent pas de l'itération choisie (traditionnelle ou bio-inspirée). Ainsi, tous nos résultats expérimentaux du chapitre 3, établis avec l'itération traditionnelle, sont valables

pour l'itération bio-inspirée. A partir de l'itération bio-inspirée, nous avons démontré quelques résultat théoriques, dont un, concernant les cycles limites en itération bio-inspirée, qui constitue l'une de nos contributions majeures. Nous pouvons déterminer, pour chaque cycle limite d'un réseau d'automates itéré en parallèle, l'ensemble des mode d'itération blocs-séquentiels (pour l'itération bio-inspirée) pour lesquels le cycle limite est présent. Ce résultat est un résultat fort pour l'itération bio-inspirée, mais également pour l'itération traditionnelle. Il nous permet notamment de retrouver, dans certains cas, comme celui du réseau de la morphogénèse florale chez *Arabidopsis thaliana*, l'ensemble des modes d'itération blocs-séquentiels (pour l'itération traditionnelle) pour lesquels le RABS modélisant le réseau de régulation génétique a des cycles limites, simplement à partir des cycles limites observés lorsque le réseau est itéré en parallèle.

Notre travail a donc contribué à montrer que la robustesse aux modes d'itération est d'une grande importance pour les RABS.

## Perspectives

A l'issue de ce travail de thèse, des perspectives intéressantes s'ouvrent.

Nos résultats expérimentaux ont montré que la robustesse des "grands" réseaux à faible connectivité était plus importante que celle des réseaux à connectivité importante. Les réseaux de régulation génétique étant des réseaux de faible connectivité et de taille relativement importante, il nous paraît intéressant de continuer à travailler dans cette voie. Tout d'abord, il serait intéressant de pouvoir générer et simuler des réseaux de taille supérieure à 7, mais de faible connectivité. Il reste, pour cela, quelques problèmes à résoudre. Le premier concerne la structure de la population des RABS de taille importante, mais de faible connectivité. Nous avons vu que, pour énumérer les RABS, et donc déterminer la structure de la population de ces réseaux, nous pouvions soit utiliser la formule de Burnside, qui nous donne un résultat exact, soit utiliser deux approximations simples à calculer. Dans le cas de réseaux de grande taille, nous avons vu que l'utilisation brute de la formule de Burnside n'était pas possible, puisqu'elle nous obligeait à faire des calculs trop coûteux en temps d'exécution. Or, dans le cas de réseaux de grande taille et de faible connectivité, les approximations sont de moins en moins bonnes à mesure que la taille augmente. Il nous faudrait donc établir de nouveaux résultats théoriques à partir de la formule de Burnside, nous permettant de réduire les temps de calcul, par exemple, en déterminant, de façon analytique, le nombre de classes d'équivalence que l'on peut construire dans des conditions particulières. Ces résultats nous permettrait de faire l'énumération exacte des RABS de grande taille, et donc de connaître la structure exacte de la population de ces réseaux. Le deuxième problème auquel nous serions confrontés, pour simuler la dynamique de grands réseaux, vient du fait que le nombre d'états possibles pour les réseaux, de même que le nombre de modes d'itération blocs-séquentiels croissent de façon exponentielle avec la taille des réseaux. En l'état actuel de nos programmes, nous estimons que l'on ne peut pas simuler,

avec des temps d'exécution raisonnables, des réseaux dont la taille est supérieure à 9 ou 10 (se posent également des problèmes de place mémoire disponible). Grâce aux résultats théoriques du chapitre 4, nous pouvons cependant réduire le nombre de modes d'itération blocs-séquentiels avec lesquels itérer un réseau, afin de déterminer la classe de comportement dynamique à laquelle il appartient. Enfin, dans le but d'avoir un modèle le plus proche possible de la réalité des réseaux de régulation génétique, il nous faudrait faire évoluer la manière dont nous construisons les RABS. On pourrait, par exemple, construire l'ensemble des réseaux de taille donnée que l'on peut obtenir en utilisant des modules de quelques noeuds (3 par exemple), connus comme étant des motifs récurrents dans les réseaux de régulation génétique, que l'on connecterait entre eux, au moyen de noeuds intermédiaires de faible degré.

Nous pensons également qu'il est important de comprendre comment les cycles limites apparaissent et disparaissent avec des changements de modes d'itération, pour les réseaux de la classe de comportement dynamique *Ev*. Nous souhaiterions prendre en compte l'ordre dans lequel les éléments sont mis à jour, et tenter de voir si des règles combinant la (dé)synchronisation des éléments d'un réseau et l'ordre de ces éléments permettraient d'expliquer, à elles seules, l'apparition et la disparition des cycles limites.

Concernant les résultats théoriques du chapitre 4, nous aimerions suivre deux pistes distinctes. La première piste serait spécifique aux RABS et consisterait à lier les propriétés des cycles limites observés pour le mode parallèle à la matrice d'interaction et au vecteur seuil du réseau. Nos résultats actuels sont basés sur les cycles limites que l'on observe en parallèle, mais ne donnent aucune indication sur les caractéristiques des réseaux qui permettent d'observer de tels cycles limites. La deuxième voie dans laquelle nous souhaiterions nous engager concerne une généralisation de nos résultats. Nos résultats actuels déterminent, pour chaque cycle limite observé pour le mode parallèle, les modes d'itération blocs-séquentiels (pour l'itération bio-inspirée) pour lesquels on observe le même cycle limite réduit. Concernant les autres modes d'itération blocs-séquentiels, nous aimerions pouvoir dire s'il y a ou non des cycles limites. Nous souhaiterions, enfin, donner des résultats à partir des cycles limites réduits observés pour d'autres modes d'itération blocs-séquentiels que le mode parallèle.

Enfin, nous avons montré que le choix du mode d'itération était, très souvent, déterminant pour la dynamique des RABS, et plus particulièrement, pour les attracteurs de ces réseaux. Les RABS se sont donc montrés peu robustes aux modes d'itération, surtout lorsque leur taille augmente. Nous avons étudié cette robustesse des RABS dans le but de mieux comprendre ce modèle de réseaux de régulation génétique. Dans le cadre de la modélisation des réseaux de régulation génétique, les résultats que nous avons obtenus sont très importants. En effet, soit cette forte sensibilité aux modes d'itération est inhérente au modèle et ne correspond pas à la réalité biologique, soit cette sensibilité peut être interprétée de façon biologique. Elle peut indiquer, par exemple, soit que la régulation géné-

tique comprend d'autres processus que ceux pris en compte dans notre modèle (et qui en assureraient la robustesse), soit que les réseaux de régulation génétique forment un échantillon très biaisé des 4 classes de comportement (en privilégiant les réseaux les plus robustes). Dans le premier cas, nous pensons qu'il est naturel et sain d'admettre que l'utilisation des RABS pour modéliser les réseaux de régulation génétique n'est pas appropriée. Il nous semble donc primordial de mettre en place une collaboration avec des biologistes, spécialistes de la régulation génétique, afin d'établir si la forte sensibilité observée des RABS aux modes d'itération, correspond effectivement à un phénomène lié à la régulation génétique. Les premiers résultats prometteurs que nous avons obtenus, concernant les *miARN*, semblent indiquer que d'autres niveaux de contrôle de la régulation génétique rendent les réseaux de régulation génétique plus robustes. Nous souhaiterions, bien entendu, continuer nos recherches dans ce sens, notamment, en modélisant de façon plus fine l'inhibition exercée par les *miARN* sur les gènes d'un réseau.

# Bibliographie


[1] G. Mendel, "Versuche über Plflanzenhybriden," *Verhandlungen des naturforschenden Vereines in Brünn*, vol. 4, pp. 3–47, 1866.

[2] H. de Vries, "Intracelluläre pangenesis," 1889. Traduction anglaise : Intracellular pangenesis. The Open Court Publishing Co., 1910.

[3] W. Johannsen, *Elemente der exakten Erblichkeitslehre*. Gustav Fischer, Jena, 1909.

[4] T. H. Morgan, "Sex limited inheritance in Drosophila," *Science*, vol. 32, pp. 120–122, 1910.

[5] G. W. Beadle and E. L. Tatum, "Genetic control of developmental reactions," *The American Naturalist*, vol. 75, p. 107, 1941.

[6] J. D. Watson and F. H. C. Crick, "A structure of deoxyribose nucleic acid," *Nature*, vol. 171, pp. 737–738, 1953.

[7] F. Jacob and J. Monod, "Genetic regulatory mechanisms in the synthesis of proteins," *Journal of Molecular Biology*, vol. 3, pp. 318–356, 1961.

[8] K. Scherrer and J. Jost, "The gene and the genon concept : a functional and information-theoretic analysis," *Molecular Systems Biology*, vol. 3, p. 87, 2007.

[9] M. Schena, D. Shalon, R. W. Davis, and P. O. Brown, "Quantitative monitoring of gene expression patterns with a complementary DNA microarray," *Science*, vol. 270, pp. 467–470, 1995.

[10] P. O. Brown and D. Botstein, "Exploring the new world of the genome with DNA microarrays," *Nature Genetics*, vol. 21, pp. 33–37, 1999.

[11] R. J. Lipshutz, S. P. Fodor, T. R. Gingeras, and D. J. Lockhart, "High density synthetic oligonucleotide arrays," *Nature Genetics*, vol. 21, pp. 20–24, 1999.

[12] D. J. Lockhart and E. A. Winzeler, "Genomics, gene expression and DNA arrays," *Nature*, vol. 405, pp. 827–836, 2000.

[13] H. de Jong, "Modeling and simulation of genetic regulatory systems : A literature review," *Journal of Computational Biology*, vol. 9, pp. 76–103, 2002.

[14] V. Volterra, "Variations and fluctuations of the number of individuals in animal species living together," *ICES Journal of Marine Science*, 1928.

[15] V. Volterra, "Variazioni e fluttuazioni del numero d'individui in specie animali conviventi," *Memorie della Accademia dei Lincei III*, vol. 6, pp. 31–113, 1926.

[16] V. Volterra, *Leçons sur la théorie mathématique de la lutte pour la vie.* Gauthier-Villars, Paris, 1931.

[17] M. Delbrück, "Discussion in : Unités biologiques douées de continuité génétique," *Colloque International du CNRS*, vol. 8, pp. 33–35, 1949.

[18] J. Demongeot, J. Aracena, S. B. Lamine, S. Meignen, A. Tonnelier, and R. Thomas, "Dynamical systems and biological regulations," in *Complex Systems*, pp. 105–149, Kluwer Academic Publishers, 2001.

[19] M. Cosnard and J. Demongeot, "On the definitions of attractors," in *Iteration Theory and its Functional Equations*, vol. 1163 of *Lecture Notes in Mathematics*, pp. 23–31, Springer, 1985.

[20] W. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Journal of Mathematical Biology*, vol. 5, pp. 115–133, 1943.

[21] S. Kauffman, "Metabolic stability and epigenesis in randomly constructed genetic nets," *Journal of Theoretical Biology*, vol. 22, pp. 437–467, 1969.

[22] R. Thomas, "Boolean formalisation of genetic control circuits," *Journal of Theoretical Biology*, vol. 42, pp. 563–585, 1973.

[23] P. Erdös and A. Rényi, "On random graphs," *Publicationes Mathematicae Debrecen*, vol. 6, pp. 290–297, 1959.

[24] P. Erdös and A. Rényi, "On the evolution of random graphs," *Publications of Mathematical Institute of Hungarian Academy of Science*, vol. 5, pp. 17–61, 1960.

[25] S. Sené, *Influence des conditions de bord dans les réseaux d'automates booléens à seuil et application à la biologie.* PhD thesis, Université Joseph Fourier de Grenoble, 2008.

[26] N. Fatès, *Robustesse de la dynamique des systèmes discrets : le cas de l'asynchronisme dans les automates cellulaires.* PhD thesis, Ecole Normale Supérieure de Lyon, 2004.

[27] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Reviews of Modern Physics*, vol. 74, pp. 47–97, 2002.

[28] S. Kauffman, *The Origins of Order : Self-Organization and Selection in Evolution.* Oxford University Press, 1993.

[29] S. Kirkpatrick, C. D. G. Jr., and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671–680, 1983.

[30] J. Holland, *Adaptation in natural and artificial systems.* University of Michigan Press, 1975.

[31] S. Liang, S. Fuhrman, and R. Somogyi, "Reveal, a general reverse engineering algorithm for inference of genetic network architecture," *Pacific Symposium on Biocomputing*, vol. 3, pp. 18–29, 1998.

[32] T. Akutsu, S. Miyano, and S. Kuhara, "Identification of genetic networks from a small number of gene expression patterns under the boolean network model," *Pacific Symposium on Biocomputing*, vol. 4, pp. 17–28, 1999.

[33] T. Akutsu, S. Miyano, and S. Kuhara, "Algorithms for inferring qualitative models of biological networks," *Pacific Symposium on Biocomputing*, vol. 5, pp. 290–301, 2000.

[34] T. Akutsu, S. Miyano, and S. Kuhara, "Inferring qualitative relations in genetic networks and metabolic pathways," *Bioinformatics*, vol. 16, pp. 727–34, 2000.

[35] J. Aracena and J. Demongeot, "Mathematical methods for inferring regulatory networks interactions : application to genetic regulation," *Acta Biotheoretica*, vol. 52, pp. 391–400, 2004.

[36] R. Thomas and R. d'Ari, *Biological Feedback*. CRC Press, 1990.

[37] R. Thomas, "Regulatory networks seen as asynchronous automata : A logical description," *Journal of Theoretical Biology*, vol. 153, pp. 1–23, 1991.

[38] R. Thomas, D. Thieffry, and M. Kaufman, "Dynamical behaviour of biological regulatory networks - I. Biological role of feedback loops and practical use of the concept of the loop-characteristic state," *Bulletin of Mathematical Biology*, vol. 57, pp. 247–276, 1995.

[39] S. Harris, B. Sawhill, A. Wuensche, and S. A. Kauffman, "A model of transcriptional regulatory networks based on biases in the observed regulation rules," *Complexity*, vol. 7, p. 23–40, 2002.

[40] S. Kauffman, C. Peterson, B. Samuelsson, and C. Troein, "Random Boolean network models and the yeast transcriptional network," *Proc Natl Acad Sci U S A*, vol. 100, pp. 14796–14799, 2003.

[41] S. Kauffman, C. Peterson, B. Samuelsson, and C. Troein, "Genetic networks with canalyzing Boolean rules are always stable," *Proc Natl Acad Sci U S A*, vol. 101, no. 49, pp. 17102–17107, 2004.

[42] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proceedings of the National Academy of Sciences of the USA*, vol. 79, pp. 2554–2558, 1982.

[43] J. J. Hopfield, "Neurons with graded response have collective computational properties like those of two-state neurons," *Proceedings of the National Academy of Sciences of the USA*, vol. 81, pp. 3088–3092, 1984.

[44] L. Mendoza and E. R. Alvarez-Buylia, "Dynamics of the genetic regulatory network for Arabidopsis thaliana flower morphogenesis," *Journal of Theoretical Biology*, vol. 193, pp. 307–319, 1998.

[45] J. Demongeot, J. Aracena, F. Thuderoz, T. Baum, and O. Cohen, "Genetic regulation networks : circuits, regulons and attractors," *Comptes Rendus de l'Académie des Sciences Biologies*, vol. 326, pp. 171–188, 2003.

[46] F. Michon, L. Forest, E. Collomb, J. Demongeot, and D. Dhouailly, "BMP2 and BMP7 play antagonistic roles in feather induction," *Development*, vol. 135, pp. 2797–2805, 2008.

[47] H. Ben-Amor, S. Cadau, A. Elena, D. Dhouailly, and J. Demongeot, "Regulatory networks analysis : robustness in morphogenesis regulation," in *AINA' 09 & BLSMC' 09*, IEEE Proceedings, Piscataway, 2009. (accepté).

[48] J. Demongeot, A. Elena, and S. Sené, "Robustness in regulatory networks : a multi-disciplinary approach," *Acta Biotheoretica*, vol. 56, pp. 27–49, 2008.

[49] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network motifs : simple building blocks of complex networks," *Science*, vol. 298, pp. 824–827, 2002.

[50] U. Alon, "Network motifs : theory and experimental approaches," *Nature Reviews Genetics*, vol. 8, pp. 450–461, 2007.

[51] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, "Network motifs in the transcriptional regulation network of Escherichia coli," *Nature Genetics*, vol. 31, pp. 64–68, 2002.

[52] S. Mangan and U. Alon, "Structure and function of the feed-forward loop network motif," *Proceedings of the National Academy of Sciences of the USA*, vol. 100, pp. 11980–11985, 2003.

[53] J. Aracena, S. B. Lamine, M. A. Mermet, O. Cohen, and J. Demongeot, "Mathematical modeling in genetic networks : relationships between the genetic expression and both chromosomic breakage and positive circuits," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 33, pp. 825–34, 2003.

[54] S. Bilke and F. Sjunnesson, "Stability of the Kauffman model," *Physical Review E*, vol. 65, p. 016129, 2002.

[55] B. Samuelsson and C. Troein, "Superpolynomial growth in the number of attractors in Kauffman networks," *Physical Review Letters*, vol. 90, p. 098701, 2003.

[56] J. E. S. Socolar and S. A. Kauffman, "Scaling in ordered and critical random boolean networks," *Physical Review Letters*, vol. 90, p. 068702, 2003.

[57] V. Kaufman, T. Mihaljev, and B. Drossel, "Scaling in critical random Boolean networks," *Physical Review E*, vol. 72, p. 046124, 2005.

[58] R. Thomas, "On the relation between the logical structure of systems and their ability to generate multiple steady states or sustained oscillations," *Springer Series in Synergetics*, vol. 9, 1981.

[59] E. Plathe, T. Mestl, and S. W. Omholt, "Feedback loops, stability and multistationarity in dynamical systems," *Journal of Biological Systems*, vol. 3, pp. 569–577, 1995.

[60] J. J. Gouzé, "Positive and negative circuits in dynamical systems," *Journal of Biological Systems*, vol. 6, pp. 11–15, 1998.

[61] E. Plathe, T. Mestl, and S. W. Omholt, "Necessary conditions for multistationarity and stable periodicity," *Journal of Biological Systems*, vol. 6, pp. 3–9, 1998.

[62] O. Cinquin and J. Demongeot, "Positive and negative feedback : striking a balance between necessary antagonists," *Journal of Theoretical Biology*, vol. 216, pp. 229–241, 2002.

[63] C. Soulé, "Mathematical approaches to differentiation and gene regulation," *Comptes Rendus de l'Académie des Sciences Biologies*, vol. 329, pp. 13–20, 2006.

[64] M. Kaufman, C. Soulé, and R. Thomas, "A new necessary condition on interaction graphs for multistationarity," *Journal of Theoretical Biology*, vol. 248, pp. 675–685, 2007.

[65] E. Remy, P. Ruet, and D. Thieffry, "Graphic requirements for multistability and attractive cycles in a Boolean dynamical framework," *Advances in Applied Mathematics*, vol. 41, pp. 335–350, 2008.

[66] A. Richard and J. P. Comet, "Necessary conditions for multistationarity in discrete dynamical systems," *Discrete Applied Mathematics*, vol. 155, pp. 2403–2413, 2007.

[67] J. Aracena, J. Demongeot, and E. Goles, "Fixed points and maximal independent sets in AND-OR networks," *Discrete Applied Mathematics*, vol. 138, pp. 277 – 288, 2004.

[68] J. Aracena, "Maximum number of fixed points in regulatory Boolean networks," *Bulletin of Mathematical Biology*, vol. 70, pp. 1398–1409, 2008.

[69] N. J. A. Sloane, "The On-line Encyclopedia of Integer Sequences : Sequence A000670." `http://www.research.att.com/~njas/sequences/A000670`, 2008.

[70] H. S. Wilf, *Generatingfunctionology.* Academic Press, NY, 1990.

[71] F. Robert, *Discrete iterations : a metric study.* Springler-Verlag, 1986.

[72] E. Goles and S. Martinez, *Neural and automata networks : dynamical behavior and applications.* Kluwer Academic Publishers, 1990.

[73] R. Thom, *Stabilité structurelle et morphgénèse.* W. A. Benjamin, 1972.

[74] E. Goles and J. Olivos, "Comportement périodique des fonctions à seuil binaires et application," *Discrete Applied Mathematics*, vol. 3, 1981.

[75] E. Goles, "Fixed point behaviour of threshold functions on a finite set," *SIAM Journal on Algebraic and Discrete Methods*, vol. 3, pp. 529–531, 1982.

[76] E. Goles-Chacc, F. Fogelman-Soulié, and D. Pellegrin, "Decreasing energy functions as a tool for studying threshold networks," *Discrete Applied Mathematics*, vol. 12, pp. 261–277, 1985.

[77] M. Cosnard and E. Goles, "Discrete state neural networks and energies," *Neural Networks*, vol. 10, pp. 327 – 334, 1997.

[78] J. Demongeot, A. Elena, and G. Weil, "Potential automata. Application to the genetic code III," *Comptes Rendus de l'Académie des Sciences Biologies*, vol. 329, pp. 953–962, 2006.

[79] L. Salinas, *Discrete models : structure and dynamics.* PhD thesis, Universidad de Chile, 2008.

[80] M. Tchuente, "Cycles generated by sequential iterations," *Discrete Applied Mathematics*, vol. 20, pp. 165 – 172, 1988.

[81] E. Goles and L. Salinas, "Comparison between parallel and serial dynamics of Boolean networks," *Theoretical Computer Science*, vol. 396, pp. 247–253, 2008.

[82] M. L. Dertouzos, *Threshold logic : a synthesis approach.* The M.I.T. Press, 1965.

[83] P. Lewis and C. Coates, *Threshold Logic.* John Wiley and Sons, 1967.

[84] S. Muroga, *Threshold logic and its applications.* Wiley-Interscience, 1971.

[85] S. Muroga, T. Tsuboi, and C. Baugh, "Enumeration of threshold functions of eight variables," *IEEE Transactions on Computers*, vol. 19, pp. 818–825, 1970.

[86] P. Ojha, "Enumeration of linear threshold functions from the lattice of hyperplane intersections," *IEEE Transactions on Neural Networks*, vol. 11, pp. 839–850, 2000.

[87] R. O. Winder, "Partitions of N-space by hyperplanes," *SIAM Journal on Applied Mathematics*, vol. 14, pp. 811–818, 1966.

[88] C. Gershenson, "Classification of random Boolean networks," in *Artificial Life VIII : Proceedings of the Eight International Conference on Artificial Life*, pp. 1–8, MIT Press, 2002.

[89] A. Elena, "Algorithme pour la simulation de la dynamique des réseaux de régulation génétique," Master's thesis, Université Joseph Fourrier, 2004.

[90] E. S. Coen and E. M. Meyerowitz, "The war of the whorls : genetic interactions controlling flower development," *Nature*, vol. 353, pp. 31–37, 1991.

[91] J. Demongeot, E. Goles, and S. Sené, "Loss of linearity and symmetrisation in regulatory networks," in *AINA' 09 & BLSMC' 09*, IEEE Proceedings, Piscataway, 2009. (accepté).

[92] R. Benlloch, A. Berbel, A. Serrano-Mislata, and F. Madueno, "Floral initiation and inflorescence architecture : a comparative view," *Annals of Botany*, vol. 100, pp. 659–676, 2007.

[93] J. G. Doench and P. A. Sharp, "Specificity of microRNA target selection in translational repression," *Genes and Development*, vol. 18, pp. 504–511, 2004.

[94] J. Brennecke, A. Stark, R. B. Russell, and S. M. Cohen, "Principles of microRNA-target recognition," *PLoS Biology*, vol. 3, pp. 404–418, 2005.

[95] B. P. Lewis, C. B. Burge, and D. P. Bartel, "Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets," *Cell*, vol. 120, pp. 15–20, 2005.

[96] M. Iorio, M. Ferracin, C. Liu, A. Veronese, R. Spizzo, S. Sabbioni, E. Magri, M. Pedriali, M. Fabbri, M. Campiglio, S. Ménard, J. Palazzo, A. Rosenberg, P. Musiani, S. Volinia, I. Nenci, G. Calin, P. Querzoli, M. Negrini, and C. Croce, "MicroRNA gene expression deregulation in human breast cancer," *Cancer Research*, vol. 65, pp. 7065–7070, 2005.

[97] J. Jouanneau and C. J. Larsen, "Les microARN : un "bras armé" du suppresseur de tumeur p53," *Bulletin du cancer*, vol. 94, pp. 634–635, 2007.

# Annexe

TAB. 2 – Vecteurs représentatifs des classes d'équivalence de fonctions booléennes à seuil à exactement 4 arguments et $a$ activations pour $0 \leq a \leq 2$.

| | | | |
|---|---|---|---|
| $a = 0$ | $(-1,-2,-2,-3)(-5)$ | $(-1,-2,-2,-3)(-4)$ | $(-1,-1,-2,-3)(-5)$ |
| | $(-1,-1,-2,-3)(-3)$ | $(-1,-1,-2,-2)(-5)$ | $(-1,-1,-2,-2)(-4)$ |
| | $(-1,-1,-2,-2)(-3)$ | $(-1,-1,-2,-2)(-2)$ | $(-1,-1,-1,-3)(-4)$ |
| | $(-1,-1,-1,-3)(-3)$ | $(-1,-1,-1,-2)(-4)$ | $(-1,-1,-1,-2)(-3)$ |
| | $(-1,-1,-1,-2)(-2)$ | $(-1,-1,-1,-1)(-4)$ | $(-1,-1,-1,-1)(-3)$ |
| | $(-1,-1,-1,-1)(-2)$ | $(-1,-1,-1,-1)(-1)$ | |
| $a = 1$ | $(1,-2,-2,-3)(-4)$ | $(1,-2,-2,-3)(-3)$ | $(1,-1,-2,-3)(-4)$ |
| | $(1,-1,-2,-3)(-2)$ | $(1,-1,-2,-2)(-4)$ | $(1,-1,-2,-2)(-3)$ |
| | $(1,-1,-2,-2)(-2)$ | $(1,-1,-2,-2)(-1)$ | $(1,-1,-1,-3)(-3)$ |
| | $(1,-1,-1,-3)(-2)$ | $(1,-1,-1,-2)(-3)$ | $(1,-1,-1,-2)(-2)$ |
| | $(1,-1,-1,-2)(-1)$ | $(1,-1,-1,-1)(-3)$ | $(1,-1,-1,-1)(-2)$ |
| | $(1,-1,-1,-1)(-1)$ | $(1,-1,-1,-1)(0)$ | $(2,-1,-2,-3)(-3)$ |
| | $(2,-1,-2,-3)(-2)$ | $(2,-1,-1,-3)(-3)$ | $(2,-1,-1,-3)(-1)$ |
| | $(2,-1,-1,-2)(-3)$ | $(2,-1,-1,-2)(-2)$ | $(2,-1,-1,-2)(-1)$ |
| | $(2,-1,-1,-2)(0)$ | $(2,-1,-1,-1)(-2)$ | $(2,-1,-1,-1)(-1)$ |
| | $(2,-1,-1,-1)(0)$ | $(3,-1,-2,-2)(-2)$ | $(3,-1,-2,-2)(-1)$ |
| | $(3,-1,-1,-2)(-2)$ | $(3,-1,-1,-2)(0)$ | $(3,-1,-1,-1)(-1)$ |
| | $(3,-1,-1,-1)(0)$ | | |
| $a = 2$ | $(1,1,-2,-3)(-3)$ | $(1,1,-2,-3)(-1)$ | $(1,1,-2,-2)(-3)$ |
| | $(1,1,-2,-2)(-2)$ | $(1,1,-2,-2)(-1)$ | $(1,1,-2,-2)(0)$ |
| | $(1,1,-1,-3)(-2)$ | $(1,1,-1,-3)(-1)$ | $(1,1,-1,-2)(-2)$ |
| | $(1,1,-1,-2)(-1)$ | $(1,1,-1,-2)(0)$ | $(1,1,-1,-1)(-2)$ |
| | $(1,1,-1,-1)(-1)$ | $(1,1,-1,-1)(0)$ | $(1,1,-1,-1)(1)$ |
| | $(1,2,-2,-3)(-2)$ | $(1,2,-2,-3)(-1)$ | $(1,2,-1,-3)(-2)$ |
| | $(1,2,-1,-3)(0)$ | $(1,2,-1,-2)(-2)$ | $(1,2,-1,-2)(-1)$ |
| | $(1,2,-1,-2)(0)$ | $(1,2,-1,-2)(1)$ | $(1,2,-1,-1)(-1)$ |
| | $(1,2,-1,-1)(0)$ | $(1,2,-1,-1)(1)$ | $(1,3,-2,-2)(-1)$ |
| | $(1,3,-2,-2)(0)$ | $(1,3,-1,-2)(-1)$ | $(1,3,-1,-2)(1)$ |
| | $(1,3,-1,-1)(0)$ | $(1,3,-1,-1)(1)$ | $(2,2,-1,-3)(-1)$ |
| | $(2,2,-1,-3)(0)$ | $(2,2,-1,-1)(-1)$ | $(2,2,-1,-1)(0)$ |
| | $(2,2,-1,-1)(1)$ | $(2,2,-1,-1)(2)$ | $(2,3,-1,-2)(0)$ |
| | $(2,3,-1,-2)(1)$ | $(2,3,-1,-1)(0)$ | $(2,3,-1,-1)(2)$ |

TAB. 3 – Vecteurs représentatifs des classes d'équivalence de fonctions booléennes à seuil à exactement 4 arguments et $a$ activations pour $3 \leq a \leq 4$.

| | | | | |
|---|---|---|---|---|
| $a = 3$ | $(1,1,1,-3)(-1)$ | $(1,1,1,-3)(0)$ | $(1,1,1,-2)(-1)$ | $(1,1,1,-2)(0)$ |
| | $(1,1,1,-2)(1)$ | $(1,1,1,-1)(-1)$ | $(1,1,1,-1)(0)$ | $(1,1,1,-1)(1)$ |
| | $(1,1,1,-1)(2)$ | $(1,1,2,-3)(-1)$ | $(1,1,2,-3)(1)$ | $(1,1,2,-2)(-1)$ |
| | $(1,1,2,-2)(0)$ | $(1,1,2,-2)(1)$ | $(1,1,2,-2)(2)$ | $(1,1,2,-1)(0)$ |
| | $(1,1,2,-1)(1)$ | $(1,1,2,-1)(2)$ | $(1,1,3,-2)(0)$ | $(1,1,3,-2)(2)$ |
| | $(1,1,3,-1)(1)$ | $(1,1,3,-1)(2)$ | $(1,2,2,-3)(0)$ | $(1,2,2,-3)(1)$ |
| | $(1,2,2,-1)(0)$ | $(1,2,2,-1)(1)$ | $(1,2,2,-1)(2)$ | $(1,2,2,-1)(3)$ |
| | $(1,2,3,-2)(1)$ | $(1,2,3,-2)(2)$ | $(1,2,3,-1)(1)$ | $(1,2,3,-1)(3)$ |
| | $(2,2,3,-1)(2)$ | $(2,2,3,-1)(3)$ | | |
| $a = 4$ | $(1,1,1,1)(0)$ | $(1,1,1,1)(1)$ | $(1,1,1,1)(2)$ | $(1,1,1,1)(3)$ |
| | $(1,1,1,2)(1)$ | $(1,1,1,2)(2)$ | $(1,1,1,2)(3)$ | $(1,1,1,3)(2)$ |
| | $(1,1,1,3)(3)$ | $(1,1,2,2)(1)$ | $(1,1,2,2)(2)$ | $(1,1,2,2)(3)$ |
| | $(1,1,2,2)(4)$ | $(1,1,2,3)(2)$ | $(1,1,2,3)(4)$ | $(1,2,2,3)(3)$ |
| | $(1,2,2,3)(4)$ | | | |

**Titre** Robustesse des réseaux d'automates booléens à seuil aux modes d'itération. Application à la modélisation des réseaux de régulation génétique.

**Résumé** Dans cette thèse, nous étudions l'influence d'un changement de mode d'itération sur les attracteurs d'un réseau d'automates booléens à seuil, outil mathématique discret classiquement utilisé pour modéliser les systèmes de régulation biologiques. L'objectif est de mettre en évidence l'importance du choix du mode d'itération pour la dynamique de ces réseaux, et en particulier pour les cycles limites atteints. Nous simulons tout d'abord la dynamique d'un échantillon non biaisé de réseaux, pour des tailles comprises entre un et sept nœuds. Les résultats des simulations montrent notamment que, lorsque la taille des réseaux croît, la dynamique de ces réseaux devient de plus en plus sensible au choix du mode d'itération. Nous démontrons ensuite un résultat théorique qui permet de déterminer, pour un réseau donné, l'ensemble des modes d'itération pour lesquels on observe des cycles limites, en fonction des cycles limites observés pour le mode parallèle

**Mots-clés** Bioinformatique, Modélisation, Réseaux de régulation génétique, Réseaux d'automates à seuil, Attracteurs, Cycles limites, Robustesse, Modes d'itération

---

**Title** Robustness of threshold boolean automata networks to iteration modes. Application to genetic regulation network modelling.

**Abstract** In this thesis, we study the influence of a change of iteration mode on the attractors of a threshold boolean automata network. These networks are discrete mathematical objects classicaly used to model biological regulation systems. The objective is to underline the importance of the choice of iteration mode for the dynamics of these networks especially for the observed limit cycles. First we run simulations of the dynamics of unbiased samples of networks with size varying from one node to seven nodes. The results of these simulations show that as the size of these networks increases their dynamics becomes more and more sensitive to the choice of iteration mode. Next we show a theoretical result that links the observed limit cycles for a network iterated with parallel mode and other iteration modes with limit cycles.

**Keywords** Bioinformatics, Modelling, Genetic regulation networks, Threshold automata networks, Attractors, Limit cycles, Robustness, Iteration modes