



HAL
open science

Modèles qualitatifs de réseaux génétiques : réduction de modèles et introduction d'un temps continu

Emilien Cornillon

► **To cite this version:**

Emilien Cornillon. Modèles qualitatifs de réseaux génétiques : réduction de modèles et introduction d'un temps continu. Autre [cs.OH]. Université Côte d'Azur, 2017. Français. NNT : 2017AZUR4085 . tel-01682212

HAL Id: tel-01682212

<https://theses.hal.science/tel-01682212>

Submitted on 12 Jan 2018

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.

École doctorale : STIC
Unité de recherche : Laboratoire I3S (UMR7271)

Thèse de doctorat

Présentée en vue de l'obtention du
grade de docteur en Informatique
de
l'Université Côte d'Azur

par

Emilien Cornillon

Modèles qualitatifs de réseaux génétiques : réduction de modèles et introduction d'un temps continu

Dirigée par Jean-Paul Comet

Soutenue le 13 octobre 2017

devant le jury composé de :

Mme. Marie	BEURTON-AIMAR	Maître de conférences	Université de Bordeaux	Rapportrice
M. Gilles	BERNOT	Professeur	Université de Nice	Co-encadrant
M. Jean-Paul	COMET	Professeur	Université de Nice	Directeur de thèse
M. Franck	DELAUNAY	Professeur	Université de Nice	Examineur
Mme. Francine	DIENER	Professeur	Université de Nice	Invitée
M. El Houssine	SNOUSSI	Professeur	Université de Rabat	Rapporteur
Mme. Fariza	TAHI	Maître de conférence	Université de Nice	Examinatrice

Remerciements

Ces quatre années de doctorat ont été très riches en expériences. J'ai eu l'occasion, en plus de mon travail de recherche, de rencontrer des personnes formidables, de monter de grands projets et de transmettre ma passion pour la science, autant via l'enseignement que via la vulgarisation scientifique.

Mes remerciements vont :

À mes deux rapporteurs Marie Beurton-Aimar, maître de conférences au LRI de Bordeaux et Houssine Snoussi, professeur à l'Université de Rabat pour le temps consacré à l'évaluation de ma thèse. Merci à Marie Beurton-Aimar pour sa bienveillance et à Houssine Snoussi pour les discussions constructives que j'ai eu l'occasion d'avoir avec lui.

Aux autres membres de mon jury : Fariza Tahy, maître de conférences à l'Université d'Evry, Francine Diener, professeur au LJAD et Franck Delaunay, professeur à l'IBV à Nice.

Je remercie particulièrement Franck Delaunay pour son apport précieux sur les aspects biologique autour du cycle circadien nécessaires à une application concrète de mes travaux de thèse, et Francine Diener, sans qui mon parcours universitaire et ma passion pour la bio-informatique n'auraient pas été les mêmes.

À Jean-Paul Comet, mon directeur de thèse, ainsi qu'à Gilles Bernot pour leur encadrement et les libertés qu'ils m'ont laissé prendre durant ces années de thèse.

À l'Université Côte d'Azur, l'Université de Nice Sophia Antipolis, et surtout le laboratoire I3S pour les cadres de recherche et d'enseignement exceptionnels dans lesquels j'ai pu faire cette thèse.

Je tiens également à remercier les membres de l'équipe MDSC, en particulier Adrien Richard pour les discussions que j'ai eu avec lui sur ses travaux et qui ont trouvé un écho dans ma thèse.

Merci à l'équipe pédagogique du département informatique de Polytech Nice Sophia et du département de Science de la Vie et de la Santé pour leurs nombreux conseils pédagogiques, en particulier Eric Gallesio, Fabien Hermenier et Patrick Coquillard.

Je tiens à remercier les collègues que j'ai côtoyés au cours de ces années de thèse, en particulier Benjamin, Jonathan, Ophélie, Cyrille, Ingrid, Guillaume, Maxime et Gilles pour leur bonne humeur.

Merci à Magali Richir pour son efficacité incroyable et sa bonne humeur à toute épreuve.

Merci également à ma famille pour leurs encouragements et surtout leurs relectures de mon manuscrit.

Merci à Sylvie pour son soutien sans faille tout au long de ces quatre années.

Résumé

Les méthodes formelles informatiques constituent un outil très puissant pour la modélisation des réseaux génétiques et pour l'étude de leur dynamique. La modélisation discrète de René Thomas permet à la fois de représenter judicieusement les connaissances biologiques et d'utiliser les méthodes formelles. Cependant, ces modèles présentent deux limitations principales : la combinatoire sous-jacente à l'identification des paramètres ne permet pas de traiter des réseaux de très grande taille et les aspects chronométriques ne sont pas pris en compte. Cette thèse offre deux contributions respectivement liées à ces questions.

La modélisation des réseaux génétiques commence par la sélection des entités les plus pertinentes pour la question abordée. Les réseaux obtenus restent souvent trop grands et nous cherchons donc à les réduire sans altérer les propriétés dynamiques importantes. Ici, nous définissons un cadre entièrement formel inspiré d'une technique d'Aurélien Naldi pour la suppression de variables et de seuils. Ces réductions conservent les comportements asymptotiques et permettent de prouver formellement l'équivalence asymptotique de différents modèles publiés d'un même réseau.

Pour prendre en compte les informations chronométriques cruciales dans certains systèmes (e.g. le cycle circadien), nous définissons un formalisme hybride fondé sur le formalisme de Thomas où les niveaux d'expression sont discrets, mais le temps continu. Ce cadre permet de construire un modèle abstrait de l'horloge circadienne des mammifères qui explique avec très peu de variables les propriétés de robustesse face à des changements de durées des alternances jour/nuit.

Abstract

Formal methods from computer science constitute a powerful tool for the modelling of gene networks, including the study of their dynamics. The discrete modelling of René Thomas allows for a proper representation of biological knowledge as well as for the use of formal methods. These models have two main limitations : the underlying combinatorics induced by parameter identification issues does not allow one to process large networks, and the chronometric aspects are not taken into account. This thesis offers two contributions according to these issues.

The design of gene network models begins with a selection of the most relevant entities. The resulting networks are often too large, and we show how to reduce them without altering the important dynamic properties. Here, we define a completely formal framework, inspired by a technique from Aurélien Naldi, driving the suppression of variables or thresholds. These reductions preserve the asymptotic behaviour. We formally prove the asymptotic equivalence of different published models for the same network.

In order to take into account chronometric information that are crucial in some systems (e.g. circadian cycle), we define a hybrid formalism based on the Thomas' formalism where expression levels are discrete but time is continuous. This framework allows for the construction of an abstract model of the circadian clock in mammals. The model explains with very few variables the robustness of the system when submitted to duration changes of the day/night alternation.

Table des matières

1	Introduction	9
1.1	La modélisation formelle des réseaux génétiques	10
1.1.1	Les réseaux de régulations génétiques	10
1.1.2	Modélisation par équations différentielles	12
1.1.3	équations différentielles linéaires par morceaux	14
1.1.4	Modélisation discrète : le formalisme de René Thomas	15
1.2	Contributions principales de la thèse	23
1.2.1	Réduction formelle de réseaux de régulation génétiques	24
1.2.2	Formalisme de modélisation hybride	25
2	Réduction de réseaux discrets de Thomas	27
2.1	Approche de réduction de Naldi	28
2.1.1	Intuition	28
2.1.2	Reformulation de la méthode de Naldi	31
2.1.3	Les limites de la méthode	34
2.2	Construction et formalisation du réseau réduit	35
2.3	Exemple de réduction	39
2.3.1	Suppression du seuil 1 de u	39
2.3.2	Suppression du seuil 1 de v	41
2.4	Discussion et conclusion	43

3	Préservation des propriétés par réduction	45
4	Application de la réduction au phage λ	59
4.1	Le mécanisme du switch chez le phage λ	60
4.2	Deux modèles classiques du phage	62
4.3	Réduction du modèle à 4 variables	64
4.3.1	Suppression de la variable booléenne CII	65
4.3.2	Suppression de la variable booléenne N	65
4.3.3	Suppression du seuil 2 de <i>Cro</i>	66
4.3.4	Suppression du seuil 1 de <i>CI</i>	67
4.3.5	Suppression du seuil 1 de <i>Cro</i>	68
4.4	Conclusion	69
5	Réseaux génétiques hybrides	71
5.1	Formalismes hybrides pour les réseaux génétiques	72
5.1.1	Motivations	72
5.1.2	Principaux formalismes hybrides existants pour les réseaux génétiques	74
5.2	Réseaux génétiques à célérité	77
5.3	États hybrides, ressources et quelques propriétés	82
5.4	Définition de la dynamique	84
5.4.1	Espace des transitions hybrides	84
5.4.2	Le cas des « hypercoins »	87
5.5	Lien avec le cadre discret de René Thomas	90
5.6	Conclusion	93
6	Modèle hybride circadien et ses propriétés remarquables	95
6.1	Le cycle circadien	96
6.1.1	Introduction	96

6.1.2	L'horloge moléculaire	97
6.2	Modèles et propriété remarquable	99
6.2.1	Le schéma des modèles	99
6.2.2	Propriété remarquable en dimension 2	100
6.3	Détermination des « pentes » des modèles	101
6.3.1	Glissement imposé en $(0, 0)$	102
6.3.2	Glissement imposé sur $(1, 0)$	104
6.3.3	Glissement imposé sur $(1, 1)$	104
6.3.4	Glissement imposé sur $(0, 1)$	105
6.4	Détermination des célérités des modèles	106
6.4.1	Glissement imposé sur $(0, 0)$	107
6.4.2	Glissement imposé sur $(1, 0)$	108
6.4.3	Glissement imposé sur $(1, 1)$	109
6.4.4	Glissement imposé sur $(0, 1)$	109
6.5	Conclusion	110
7	Rafinements des modèles circadiens avec CircaPlot	111
7.1	CircaPlot : un programme de visualisation des trajectoires	112
7.1.1	Motivations	112
7.1.2	Vision utilisateur	113
7.1.3	Structure du programme	114
7.1.4	Protocoles d'identification des célérités	115
7.2	Premier modèle : avec état stable	117
7.2.1	Modèle discret pour la drosophile	117
7.2.2	Modèle mammifère avec état stable	118
7.2.3	Test du modèle en obscurité constante et en équinoxe	120
7.2.4	Test de résistance du modèle aux saisons	121

7.2.5	Test de résistance du modèle aux décalages horaires	122
7.2.6	Bilan	124
7.3	Second modèle : sans état stable	125
7.3.1	Test du modèle en obscurité constante et en équinoxe	126
7.3.2	Test de résistance du modèle aux saisons	126
7.3.3	Test de résistance du modèle aux décalages horaires	127
7.4	Conclusion	130
8	Prolongements du formalisme hybride	131
8.1	plateforme de simulation HyMBioNet	132
8.1.1	Interface de description d'un réseau hybride de régulation génétique	132
8.1.2	Interface de simulation	134
8.1.3	Résultats pour le modèle de l'horloge	137
8.2	Identification des contraintes sur les célérités	138
8.2.1	Introduction à la logique de Hoare pour les réseaux génétiques . . .	139
8.2.2	Extension de la logique de Hoare au cadre de modélisation hybride .	140
8.3	Identification des contraintes sur le modèle de l'horloge	141
8.4	Identification des contraintes pour un modèle du cycle cellulaire	144
8.4.1	Le réseau du cycle cellulaire	144
8.4.2	Triplet de Hoare	145
8.4.3	Simulation	148
8.5	Conclusion	148
9	Conclusion et perspectives	151
9.1	Contributions	152
9.1.1	Réduction formelle de réseaux de régulation génétique	152
9.1.2	Cadre de modélisation hybride	153

9.2	Perspectives	154
9.2.1	Réduction symbolique	154
9.2.2	Modélisation hybride	155
9.2.3	Réduction de modèles hybrides	156

Chapitre 1

Introduction

Cette thèse se situe résolument dans le contexte de la modélisation des réseaux génétiques en suivant et en étendant l’approche proposée par René Thomas et Houssine Snoussi entre les années 70 et 90 [69, 70, 30] et enrichie depuis, en particulier par l’apport de méthodes formelles [15, 16]. Une motivation forte tout au long de cette thèse a été de favoriser la conception de modèles de petite taille dédiés à des questions précises. Nous avons abordé cette motivation essentiellement sous deux angles : réduire la taille d’un modèle en préservant des comportements limites et développer un formalisme hybride permettant de prendre en compte des informations temporelles sans nuire à l’application de méthodes formelles et sans augmenter le nombre de variables nécessaires.

Dans ce chapitre, nous commençons par présenter le minimum des notions biologiques nécessaires. Nous présentons ensuite rapidement les méthodes de modélisation par équations différentielles, puis, plus en détail, le formalisme discret de René Thomas avec multiplexes et enfin les techniques formelles d’identification des paramètres pour ce formalisme. Finalement, nous présenterons les principales contributions apportées au cours de cette thèse.

1.1 La modélisation formelle des réseaux génétiques

Les mécanismes biologiques sont généralement composés de très nombreuses entités qui interagissent et permettent l’émergence de dynamiques complexes. Cette complexité rend la compréhension de ces mécanismes très difficile. La modélisation facilite cette compréhension en permettant, d’une part, une meilleure compréhension des causalités qui pilotent la dynamique, et d’autre part, la « prédiction » de comportements non connus.

Dans cette section, nous commençons par un rappel rapide de la démarche de modélisation des processus moléculaires qui sous-tendent les réseaux génétiques, puis nous présentons les divers formalismes mathématiques classiques utilisés pour les modéliser.

1.1.1 Les réseaux de régulations génétiques

La classe de mécanismes biologiques étudiés dans cette thèse est la classe des *réseaux de régulations génétiques* (RRG). Ces réseaux sont caractérisés par la présence de régulations moléculaires de gènes (activations ou inhibitions). Sur l’ADN, la séquence associée à un gène possède une zone appelée un *promoteur*. Le promoteur est spécifique à une ou plusieurs protéines qui, en se fixant dessus, activent ou bloquent le gène. La plupart du temps, quand un gène est activé, il est transcrit en ARN messenger qui, après une série de traitements, va être traduit en protéine. Cette protéine, après maturation, va, par exemple, interagir avec d’autres entités cellulaires. Les interactions peuvent être des régulations directes de gènes, d’ARN, ubiquitination ou formation de complexes avec d’autres protéines

qui peuvent modifier les fonctions des molécules.

Contrairement aux réseaux métaboliques, un réseau génétique ne « consomme » pas de molécules : d'une part, la transcription n'altère pas la séquence du gène qui peut donc être à nouveau utilisée immédiatement, et d'autre part, les protéines qui se fixent sur le promoteur sont libérées dès le début de la transcription et peuvent donc être réutilisées pour commencer une nouvelle transcription. Un gène peut donc produire des protéines sans consommer de ressources moléculaires ni être en concurrence avec d'autres gènes. De plus, les réseaux génétiques sont des réactions autocatalytiques. Une transcription d'un gène ne s'effectuant que sur un brin d'ADN, elle nécessite l'ouverture de la double hélice. Une fois ouverte, le temps de transcription est grandement réduit. Pour cette raison, l'évolution de la vitesse de transcription d'un gène a une forme sigmoïdale lorsqu'il s'active ou se désactive.

Les réseaux de régulations génétiques comptent souvent de nombreux gènes et par conséquent au moins autant de protéines et d'ARN messenger. Leur dynamique devient alors excessivement difficile à appréhender. Pour comprendre les causalités, il est donc nécessaire de pouvoir simplifier les aspects moléculaires d'un réseau génétique afin d'en faciliter la compréhension et la modélisation. La construction d'un modèle repose donc sur un ensemble d'*hypothèses raisonnables*. La démarche de modélisation comporte quelques étapes clés :

1. Choisir les entités abstraites qui vont interagir dans le réseau. Selon la question posée, toutes les entités n'auront pas forcément le même impact sur le comportement du modèle (effet négligeable ou cible extérieure au réseau). Il en résulte que certains interacteurs moléculaires seront simplement ignorés, d'autres regroupés sous une même entité cohérente avec la question posée, *etc.* Il est important de limiter le nombre d'entités abstraites, car la difficulté à identifier les paramètres du modèle final dépendra, de manière exponentielle, de la taille du réseau génétique construit.
2. Faire un inventaire des interactions entre les entités (activations, inhibitions, coopérations en formant des complexes, interdépendances, seuils d'actions, *etc.*). Il résulte de cette étape un graphe d'interactions annoté entre les entités.
3. Encoder ce graphe selon un formalisme : si par exemple une entité e est sous l'influence des entités e_1 à e_n , le formalisme des équations différentielles fera intervenir les taux de concentration des produits de e_1 à e_n dans l'expression de la dérivée de e par rapport au temps.
4. L'encodage précédent fait toujours intervenir des paramètres symboliques dont on ne connaît pas *a priori* la valeur. Commence alors une étape d'identification des paramètres dont l'objectif est de mettre en conformité la dynamique du modèle avec les connaissances biologiques.
5. La modélisation a généralement pour but de faire des « prédictions » *in silico* :

vérification d'hypothèses (que l'on formalise à cette occasion), simuler le modèle dans des situations non testées expérimentalement, screening virtuel, proposition d'expériences pour compléter l'identification des paramètres, *etc.* (ensemble d'activités souvent regroupées sous le nom de « retour paille »).

Bien souvent, chacune de ces 5 étapes clés impose un retour vers les étapes précédentes afin d'affiner le modèle ou pour résoudre des incohérences.

Les différentes simplifications et le choix du formalisme sont, en pratique, guidés par la question biologique qui motive la modélisation. En ce qui concerne les formalismes, nous en détaillons ci-dessous trois parmi les plus courants.

1.1.2 Modélisation par équations différentielles

La modélisation par équations différentielles des réseaux génétiques est principalement fondée sur la détermination de la vitesse de changement du taux de concentration du produit de chaque gène. Il y a donc, en général, une équation différentielle par gène (avec quelques exceptions pour lesquelles il est pertinent de distinguer explicitement, par exemple, l'ARN messenger de sa protéine). Cette approche par *système d'équations différentielles* est celle qui a été longtemps majoritairement utilisée pour la modélisation des réseaux génétiques [37, 49, 74, 19], bien que cette dernière décennie ait vu une croissance remarquablement forte des approches discrètes et hybrides et en particulier celles fondées sur l'approche de René Thomas.

Les équations différentielles constituent un formalisme purement continu dans le sens où, à la fois, le temps et les taux de concentration sont des nombres réels qui varient de manière continue. Chacune des équations du système d'équations différentielles détermine à chaque instant quelle est la dérivée (vitesse d'évolution locale) du taux de concentration du gène considéré. Dans le cadre des réseaux génétiques, ces équations ont une forme généralement très standardisée :

$$\frac{dx}{dt} = k_0 + f_1(y_1) + \dots + f_n(y_n) - \gamma x$$

Le premier terme k_0 indique une production constitutive minimale du gène x lorsque celui-ci ne bénéficie d'aucun activateur et est réprimé par tous ses inhibiteurs.

Les termes suivants de la forme $f_i(y_j)$ constituent un inventaire des activateurs ou inhibiteurs du gène. Si y_i est le taux de concentration du produit du $i^{\text{ème}}$ gène, alors on reconnaît un activateur par le fait que f_i soit croissante et un inhibiteur par le fait que f_i soit décroissante.

Lorsque x est un gène, les fonctions f_i sont généralement sigmoïdales parce que la transcription d'un gène est un phénomène autocatalytique. Pour cette raison, on

utilise généralement une fonction de Hill $f_i(y_i) = \frac{y_i^n}{k_i + y_i^n}$ à une constante multiplicative près. La puissance n de la fonction de Hill contrôle la pente montante. Pour un inhibiteur, la fonction de Hill devient $f_i(y_i) = \frac{k_i}{k_i + y_i^n}$.

Le dernier terme γx indique la dégradation constitutive du produit du gène. Intuitivement, γ représente l'efficacité des protéases (si le produit de x est une protéine) et la vitesse de dégradation est proportionnelle à x puisque, par exemple, deux fois plus de produits de x dans la cellule multiplie par deux la probabilité de rencontre avec une protéase.

Dans tous les cas la constante k_0 et les fonctions f_i sont positives, que y_i soit un activateur ou un inhibiteur. Seule la dégradation a une action négative sur la dérivée de x . Ces propriétés sont raisonnables puisque, comme déjà mentionné, un gène ne consomme aucun produit.

Dans certains cas, il peut arriver que des coopérations entre les y_i se produisent et donnent lieu à des termes qui peuvent dépendre, par exemple, d'une somme $f_{ij}(y_i + y_j)$ ou d'un produit $f_{ij}(y_i y_j)$.

Puisque les équations différentielles sont fondées sur des taux de concentration réels, elles font l'hypothèse d'un très grand nombre de molécules de chaque type (pour qu'il soit assimilable à une infinité de molécules). Par conséquent, si l'on souhaite modéliser des gènes dont les protéines agissent en faible nombre, les équations différentielles ne sont pas adaptées. Pour la même raison, les trajectoires qui sont solutions d'un tel système d'équations différentielles sont des trajectoires déterministes très précises.

Les équations différentielles présentent l'avantage de pouvoir exploiter des données quantitatives issues d'expériences biologiques puisque la solution d'une équation différentielle est par définition quantitative et peut donc être comparée à des données biologiques. En revanche, un système d'équations différentielles génère beaucoup de paramètres réels dont on ne connaît pas les valeurs (k_0 , les k_i et chacun des coefficients multiplicateurs des fonctions f_i , et le taux de dégradation γ). Ce n'est pas tant le nombre de paramètres inconnus qui rend leur identification difficile que le fait que ces paramètres soient à valeurs réelles. En effet, une modification mineure de la valeur d'un paramètre peut tout à fait franchir un seuil de bifurcation qui change drastiquement les trajectoires.

La plupart des modélisateurs par équations différentielles sont donc amenés à faire confiance à des valeurs le plus souvent mesurables *in vitro* plutôt qu'*in vivo* et à mener un très grand nombre de simulations pour découvrir par essai-erreur les valeurs des paramètres non mesurables ou fortement dépendants du contexte *in vivo*. Cette approche « à tâtons » ne peut que fournir un jeu de paramètres *parmi d'autres* qui produit des simulations compatibles avec les observations biologiques, de sorte que le plus souvent, une connaissance biologique supplémentaire conduit à recommencer presque *ex nihilo* le lourd travail d'identification des paramètres. De plus, il est presque impossible de déléguer entièrement cette tâche d'identification des paramètres à l'ordinateur parce qu'il est très

difficile de formaliser l'impact d'un changement de paramètre sur les trajectoires.

De nombreuses questions biologiques sont en fait de nature qualitative et des mesures approximées suffisent souvent à y répondre. Il en résulte que dans la plupart des cas, la précision offerte par les équations différentielles dépasse les besoins de modélisation nécessaires pour répondre à une question biologique donnée. Les lourds efforts d'identification des paramètres esquissés précédemment sont donc, d'une certaine façon, assez mal payés de retour. C'est pour cette raison que depuis plusieurs décennies, beaucoup de recherches tendent à proposer des formalismes simplifiés par rapport à celui des équations différentielles présenté ici.

1.1.3 équations différentielles linéaires par morceaux

Une première étape de simplification consiste à découper l'espace des concentrations en intervalles « qualitativement homogènes » et à supprimer les fonctions de Hill, en affectant à chaque portion de l'espace des équations différentielles linéaires par morceaux. Reprenons les équations générales de la forme

$$\frac{dx}{dt} = k_0 + f_1(y_1) + \dots + f_n(y_n) - \gamma x$$

La variable y_1 , par exemple, participe *a priori* à plusieurs des équations différentielles du système d'équations, correspondant à ses cibles x_1, \dots, x_k . Pour chacune d'elles, il y a une fonction de Hill dont on peut considérer le point d'inflexion. Cela munit donc y_1 d'une collection de points d'inflexion que l'on peut classer en ordre croissant $\Theta_{1,x_1}, \dots, \Theta_{1,x_k}$ et ces seuils déterminent donc une suite d'intervalles qui couvre l'espace des taux de concentration possibles de y_1 . En faisant de même pour toutes les variables du système, on découpe donc l'espace des concentrations en hyperpavés.

L'idée est de placer dans chacun de ces hyperpavés un système d'équations différentielles linéaires, puisque ces équations possèdent une solution analytique simple. Dés lors, la forme d'une équation générale devient

$$\frac{dx}{dt} = k_0 + k_1 \mathbb{1}_{\Theta_1}^{\pm}(y_1) + \dots + k_n \mathbb{1}_{\Theta_n}^{\pm}(y_n) - \gamma x$$

où les k_i sont des nombres réels positifs et où, par convention, si y_i est un activateur alors on utilise la fonction indicatrice $\mathbb{1}_{\Theta_i}^+$ qui vaut 0 si $y_i < \Theta_i$ et 1 si $y_i \geq \Theta_i$, et si y_i est un inhibiteur alors on utilise $\mathbb{1}_{\Theta_i}^-$ qui vaut 1 si $y_i < \Theta_i$ et 0 si $y_i \geq \Theta_i$ (voir Figure 1.1). Dans chacun des hyperpavés, l'équation différentielle devient donc simplement

$$\frac{dx}{dt} = S - \gamma x \quad (\text{où } S \text{ est la somme des } k_i \text{ tels que } \mathbb{1}_{\Theta_i}^{\pm} \text{ vaut } 1)$$

dont la solution est de la forme $Ce^{-\gamma t} + \frac{S}{\gamma}$, et dont la limite quand t tend vers l'infini est

$\frac{S}{\gamma}$.

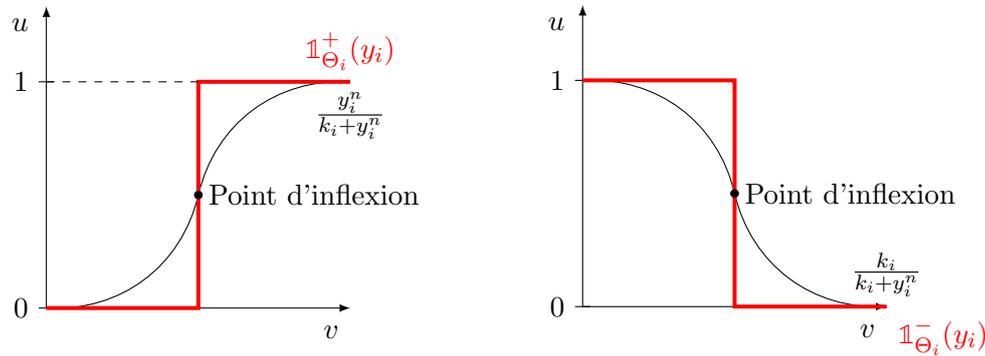


FIGURE 1.1 – Approximation d’une fonction de Hill par une fonction seuil. Θ_i est le point d’inflexion de la fonction de Hill. Dans le cadre des réseaux génétiques, l’exposant n de la fonction de Hill rend la sigmoïde correspondante rapidement très raide et par conséquent, l’approximation est très fidèle. à gauche, le cas d’une activation et à droite, le cas d’une inhibition.

Comme on le voit, les équations différentielles linéaires par morceaux simplifient le formalisme précédent où les trajectoires étaient dérivables partout : il suffit maintenant de trouver les valeurs des nombres réels k_i au lieu de trouver tous les paramètres des fonctions de Hill f_i . Il n’en reste pas moins que chacun des k_i peut prendre une infinité de valeurs réelles même si les données expérimentales peuvent souvent permettre de contraindre k_i dans un petit intervalle.

1.1.4 Modélisation discrète : le formalisme de René Thomas

Un moyen de réduire l’ensemble des valeurs possibles des paramètres est de suivre un formalisme de modélisation discrète puisque la connaissance d’un intervalle des valeurs raisonnables de chaque paramètre fournit un nombre fini de valeurs possibles. L’identification des paramètres devient plus simple.

Le formalisme discret multivalué de René Thomas et Houssine Snoussi [69, 71] reprend le principe de découpage de l’espace des concentrations du formalisme d’équations différentielles par morceaux et numérote les intervalles par ordre croissant à partir de 0. De ce fait, le numéro d’un intervalle indique le nombre de cibles sur lesquelles le gène considéré est actif. Ce découpage qualitatif de l’espace des concentrations est, la plupart du temps, en parfait accord avec le mode de raisonnement des biologistes. En effet, dans un réseau génétique, les biologistes sont généralement intéressés par les comportements qualitatifs de chacun des gènes (excès d’une protéine, inhibition ou activation d’un gène, oscillation ou stabilité).

Cette discrétisation par intervalle du formalisme de Thomas a pour conséquence de discrétiser également le temps : on ne retient en effet que les changements d’intervalles

et une trajectoire du réseau génétique est donc décrite par une succession de gènes qui changent d'intervalle. Le temps est donc, de fait, discret.

De par son abstraction, le formalisme de Thomas est plus général que la seule modélisation des réseaux génétiques. En effet, déjà dans le domaine des réseaux génétiques, un « gène » représente le regroupement du gène, de son ARN, des protéines traduites, *etc.* Pour cette raison, le formalisme de Thomas préfère à la notion de « gène » la notion de « variable » qui est plus générale. Une variable peut représenter tout aussi bien un gène qu'une condition expérimentale (comme la présence de lumière). Il peut même être utilisé le formalisme de Thomas pour des modèles non génétiques très généraux [45].

Définition d'un réseau génétique

La définition du formalisme que nous donnons ici est une extension de celui présenté par Thomas et Snoussi, à laquelle a été ajoutée la notion de multiplexes [44]. Les multiplexes sont composés par un symbole associé à une formule. Ils sont utilisés pour spécifier et encoder en logique propositionnelle les comportements des interactions. Les informations de ce type étaient contenues dans les valeurs de paramètres alors qu'avec les multiplexes, il est possible de mettre cette information dans les formules, ce qui, d'une part, augmente grandement la lisibilité du graphe d'interaction, et d'autre part, diminue drastiquement le nombre de paramètres du modèle puisqu'une coopération *via* un multiplexe factorise les interactions. La définition 1 définit le langage des formules des multiplexes.

Définition 1. [*Langage des multiplexes*] Soit V un ensemble de variables tel que chaque variable $v \in V$ est muni d'une borne $b_v \in \mathbb{N}^*$. Le langage \mathcal{L}_V est défini inductivement par :

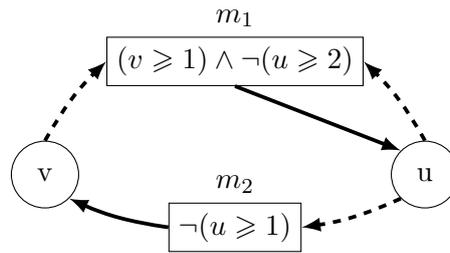
- Si $v \in V$ et n est un entier tel que $1 \leq n \leq b_v$, alors $v \geq n$ est un atome de \mathcal{L}_V
- Si φ et ψ sont deux formules de \mathcal{L}_V , alors $\neg\varphi$, $(\varphi \vee \psi)$, $(\varphi \wedge \psi)$ et $(\varphi \Rightarrow \psi)$ appartiennent aussi à \mathcal{L}_V

Dans les cas simples d'une activation ou d'une inhibition d'un gène par un autre gène, la formule du multiplexe correspondant ne comportera qu'un seul atome : pour une variable v , la formule sera $(v \geq n)$ dans le cas d'une activation au seuil n et $\neg(v \geq n)$ dans le cas d'une inhibition. De plus, l'utilisation de la logique permet de décrire des interactions plus complexes. Par exemple, prenons un complexe de deux protéines a et b , activant un gène v . Plutôt que de poser deux multiplexes différents, on représentera cette formation de complexe par un multiplexe contenant la conjonction de 2 atomes $(a \geq p) \wedge (b \geq q)$. Cette formule indique qu'il faut que a et b dépassent tous deux leurs seuils respectifs pour que la formule soit vraie, c'est-à-dire pour que le complexe A-B soit actif.

Définition 2. [*Réseau de régulation génétique*] Un réseau de régulation génétique (RRG) est un quadruplet $R = (V, M, E, \mathcal{K})$ où :

- V est un ensemble fini de variables dans lequel chaque $v \in V$ est muni d'une borne $b_v \in \mathbb{N}^*$.
- M est un ensemble fini de multiplexes dans lequel chaque symbole $m \in M$ est muni d'une formule φ_m appartenant au langage \mathcal{L}_V .
- E est un ensemble d'arcs de la forme $(m \rightarrow v) \in M \times V$.
- $\mathcal{K} = \{K_{v,\omega}\}$ est une famille d'entiers tels que $0 \leq K_{v,\omega} \leq b_v$, indexées par l'ensemble des couples (v,ω) où $v \in V$ et ω est un sous-ensemble quelconque de $R^-(v)$ où $R^-(v) = \{m \in M \mid (m \rightarrow v) \in E\}$, autrement dit, ω est un ensemble de prédécesseurs de v dans le réseau.
 $K_{v,\omega}$ est appelé le paramètre de v pour l'ensemble de ressources ω .

La figure 1.2 contient un exemple de graphe d'interactions et le tableau en dessous du graphe détaille les éléments du quadruplet correspondant. Dans cet exemple, les valeurs données aux paramètres K sont complètement arbitraires. Nous verrons plus loin comment ces paramètres influencent la dynamique.



$V :$	$M :$	$E :$	$\mathcal{K} :$
$v (b_v = 1)$	$m_1 : (v \geq 1) \wedge \neg(u \geq 1)$	$(m_1 \rightarrow u)$	$K_{v,\emptyset} = 0$ $K_{u,\emptyset} = 1$
$u (b_u = 2)$	$m_2 : \neg(u \geq 1)$	$(m_2 \rightarrow v)$	$K_{v,\{m_2\}} = 1$ $K_{u,\{m_1\}} = 2$

FIGURE 1.2 – Exemple de réseau de régulation génétique. En haut, le graphe d'interaction et en bas, le quadruplet du réseau. La formule $\neg(u \geq 1)$ du multiplexe m_2 signifie que u inhibe v ; ce qui peut également se dire « m_2 est une ressource de v quand u est absent ». La formule $(v \geq 1) \wedge \neg(u \geq 2)$ signifie que m_1 est une ressource de u quand v est actif et que u lui-même n'est pas « surexprimé ».

Dynamique d'un modèle

Le formalisme de Thomas est un formalisme discret tel que chaque modèle a un nombre fini d'états. Ils correspondent à toutes les combinaisons de valeurs que peut prendre chacune des variables. Dans l'exemple de la figure 1.2, la variable u a 3 valeurs possibles : 0 (inactif), 1 (actif sur v) et 2 (actif sur v et sur lui-même). De même, la variable v a 2 valeurs

possibles : 0 (inactif), 1 (actif sur u). Il y a donc 6 états :

				$v \uparrow$
	1	(0,1)	(1,1)	(2,1)
	0	(0,0)	(1,0)	(2,0)
		0	1	2 $u \rightarrow$

La définition suivante formalise cette notion d'états discrets.

Définition 3. [état discret] Soit $R = (V, M, E, \mathcal{K})$ un RRG. Un état discret η de R est une fonction $\eta : V \rightarrow \mathbb{N}$ telle que, pour tout $v \in V$, $\eta(v) \leq b_v$.

Par abus de notation et quand le contexte donne l'ordre des variables, pour un ensemble des variables $V = \{v_1, \dots, v_n\}$, nous noterons (x_1, \dots, x_n) l'état tel que $x_i = \eta(v_i)$.

Dans chaque état discret, chacune des variables du réseau est susceptible d'être influencée par les autres au travers d'un ou plusieurs multiplexes. On dit qu'un multiplexe est ressource d'une variable dans un état discret donné lorsque cet état satisfait la formule du multiplexe. La définition suivante introduit formellement la notion de ressource.

Définition 4. [Ressources] Soit $R = (V, M, E, \mathcal{K})$ un RRG et $v \in V$ une variable de R . La relation de satisfaction $\eta \models \varphi$ (où η est un état discret de R et φ une formule de \mathcal{L}_V) est inductivement définie par :

- Si φ est l'atome $v \geq n$, alors $\eta \models \varphi$ si et seulement si $\eta(v) \geq n$
- Si φ est de la forme $\neg\psi$, alors $\eta \models \varphi$ si et seulement si $\eta \not\models \psi$
- Si φ est de la forme $\psi_1 \vee \psi_2$, alors $\eta \models \varphi$ si et seulement si $\eta \models \psi_1$ ou $\eta \models \psi_2$
- Si φ est de la forme $\psi_1 \wedge \psi_2$, alors $\eta \models \varphi$ si et seulement si $\eta \models \neg(\neg\psi_1 \vee \neg\psi_2)$
- Si φ est de la forme $\psi_1 \Rightarrow \psi_2$, alors $\eta \models \varphi$ si et seulement si $\eta \models (\psi_2 \vee \neg\psi_1)$.

L'ensemble des ressources de la variable v pour l'état η est défini par : $\rho(\eta, v) = \{m \in R^-(v) \mid \eta \models \varphi_m\}$, c'est-à-dire l'ensemble de tous les multiplexes prédécesseurs de v pour lesquels la formule est satisfaite.

Comme nous l'avons vu précédemment, les multiplexes permettent de regrouper des informations. Ce regroupement a pour effet de réduire le nombre paramètres codant le comportement de la variable. Si chacune des variables agissant sur une cible possède leur propre multiplexe, le nombre de paramètres est de 2^i où i est le nombre de prédécesseurs de la cible. Le regroupement d'interactions dans des multiplexes peut alors grandement diminuer ce nombre. Par exemple, reprenons le cas d'une coopération de deux gènes a et b dont les protéines forment un complexe qui active un troisième gène v . La figure 1.3 contient les deux versions du réseau. à gauche, la version « atomique » résultant du formalisme classique de Thomas, dans laquelle chaque influence sur v est indépendante, et à droite la version qui utilise la puissance des multiplexes. Supposons que dans la figure de droite $K_{v,\emptyset} = 0$ et $K_{v,\{m_{AB}\}} = 1$, alors on peut encoder dans la figure de gauche le multiplexe m_{AB} avec les valeurs de paramètres suivantes $K_{v,\emptyset} = K_{v,\{m_A\}} = K_{v,\{m_B\}} = 0$ et $K_{v,\{m_A, m_B\}} = 1$. De même, on aurait pu encoder une disjonction ou toute autre formule.

Les multiplexes sont donc un moyen d'exploiter une connaissance biologique pour réduire le nombre de paramètres à identifier.

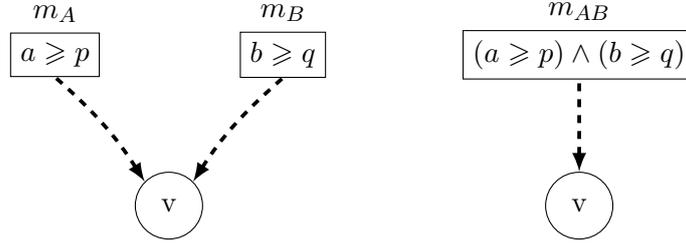


FIGURE 1.3 – Exemple de réduction du nombre de ressources en utilisant les multiplexes. à gauche, un graphe d'interactions représentant un graphe de Thomas et à droite le graphe d'interactions après fusion des multiplexes. à gauche, 4 paramètres sont nécessaires : $K_{v,\emptyset}$, $K_{v,\{m_A\}}$, $K_{v,\{m_B\}}$ et $K_{v,\{m_A,m_B\}}$ permettant d'encoder tous les cas de comportement (conjonction de m_A et m_B , mais aussi disjonction, ou exclusif ...). à droite, en exploitant la connaissance que le complexe A-B implique une conjonction, il ne reste plus que 2 paramètres : $K_{v,\emptyset}$ et $K_{v,\{m_{AB}\}}$.

Remarquons que, comme dans la théorie initiale de René Thomas, les paramètres K peuvent rendre une interaction non fonctionnelle. En effet, dans la figure 1.3 droite, si l'on choisit $K_{v,\emptyset} = K_{v,\{m_{AB}\}} = 0$ ou $K_{v,\emptyset} = K_{v,\{m_{AB}\}} = 1$ alors on rend l'action de m_{AB} sur v non fonctionnelle car, comme on le verra plus tard, le comportement de v sera piloté par la même valeur de paramètres que la formule de m_{AB} soit vraie ou fausse.

La figure 1.4 contient un exemple qui réduit encore plus le nombre de paramètres à identifier : si un ensemble de ressources ω contient deux multiplexes m_1 et m_2 tel que la conjonction φ_{m_1} et φ_{m_2} est non satisfiable (il n'existe pas d'états tels que ces deux formules soient vraies) alors $K_{v,\omega}$ peut être ignoré puisqu'il ne sera utilisé dans aucun état. Plus généralement, si $(\bigwedge_{m \in \omega} \varphi_m)$ est non satisfiable alors $K_{v,\omega}$ peut être ignoré.

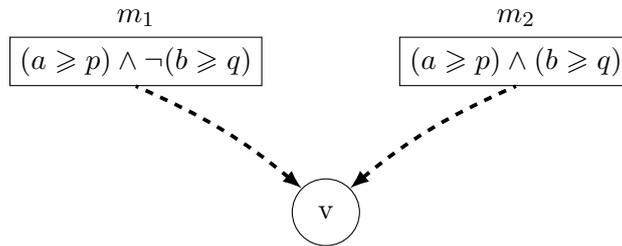


FIGURE 1.4 – Exemple de deux multiplexes ne pouvant pas être ressource de v simultanément. $K_{v,\{m_1,m_2\}}$ ne sera jamais utilisé dans la dynamique.

Du point de vue de la dynamique du système, dans un état donné, si l'ensemble des ressources de v est ω , le paramètre $K_{v,\omega}$ est la valeur vers laquelle v est attirée. C'est sa *valeur focale*. Plus généralement, le *point focal* d'un état discret η est l'état dans lequel les composantes de toutes les variables ont comme valeur la valeur focale à l'état η , autrement dit, l'état η' tel que, pour tout $v \in V$, $\eta'(v) = K_{v,\rho(\eta,v)}$. Cependant, le système n'ira pas

forcément de l'état η à η' car il est soumis aux deux idées clés de Thomas suivantes :

Désynchronisation Quand le point focal d'un état fait évoluer au moins deux variables en même temps, par exemple vers une augmentation de ces deux variables, cela signifie que les gènes correspondants synthétisent leurs protéines plus rapidement qu'elles ne sont dégradées. Le prochain seuil à franchir pour chacun de ces gènes correspond à un nombre de protéines fixé par le seuil. Par conséquent, la probabilité de traverser *simultanément* les seuils de ces gènes est nulle. Nous ne pouvons cependant pas savoir lequel des deux seuils sera traversé en premier. Il faut donc modéliser la possibilité de changement de chacune de ces variables de manière non déterministe. Intuitivement, cette désynchronisation fait que la différence entre avant et après le seuil est la présence d'une molécule supplémentaire : la probabilité que deux molécules soient produites exactement au même moment, et en nombres à la limite des seuils respectifs, est nulle. Plus généralement, ce raisonnement s'applique à un nombre quelconque de gènes et on n'autorisera qu'un seul passage de seuil à la fois.

Transitions unitaires Quand la valeur focale d'une variable, dans un état donné, l'oblige à traverser plusieurs états pour l'atteindre, l'augmentation ou la diminution progressive du nombre de protéines fait que le taux de concentration de la variable ne « sautera » pas directement de l'état courant à la valeur focale : il passera d'abord dans l'état voisin de l'état courant. Chacune des transitions est donc de longueur 1 dans la direction de la valeur focale : un paramètre plus grand que la valeur courante donnera une transition faisant croître l'état de la variable et un paramètre plus petit, une transition faisant décroître l'état de la variable.

Les ressources de l'état voisin atteint selon ces deux principes, et donc selon ses valeurs focales, ne sont pas nécessairement les mêmes que dans l'état initial. Il en résulte que la valeur focale initiale peut éventuellement ne jamais être atteinte à partir de l'état initial.

La définition suivante reflète ces deux idées clés :

Définition 5. [graphe d'états] Soit $A = (V, M, E, \mathcal{K})$ un RRG, le graphe d'états de A est un couple $\mathcal{G} = (S, T)$ où :

- S est l'ensemble des états discrets de A .
- $T \subset S \times S$ est l'ensemble des transitions de A . Il existe une transition de η vers η' , notée $\eta \rightarrow \eta'$, si et seulement si il existe une variable $v \in V$ telle que :
 - $K_{v, \rho(\eta, v)} \neq \eta(v)$ et
 - si $\eta(v) > K_{v, \rho(\eta, v)}$, alors $\eta'(v) = \eta(v) - 1$
 - si $\eta(v) < K_{v, \rho(\eta, v)}$ alors $\eta'(v) = \eta(v) + 1$
 - Pour tout $u \neq v$, nous avons $\eta'(u) = \eta(u)$.

Un chemin correspond à une succession d'états discrets connectés par des transitions qui permet d'aller à un état d'arrivée en partant d'un état d'origine. La désynchronisation des transitions faisant du graphe d'états un graphe non déterministe, elle lui permet de

contenir l'ensemble exhaustif des chemins possibles dans la dynamique du réseau. Un « modèle » est alors un modèle logique dans lequel les paramètres codent pour une dynamique (ensemble de chemins) précise. Un changement de paramètres qui entraîne un changement du graphe d'états génère donc un autre modèle.

Pour illustrer cette définition, reprenons le graphe d'interaction de la figure 1.2. La figure 1.5 contient des valeurs de paramètres choisies arbitrairement et, en partie basse, les valeurs focales de chacune des variables pour chacun état discret, compte tenu de ces valeurs de paramètres.

Paramètres de v		Paramètres de u	
$K_{v,\emptyset} = 0$		$K_{u,\emptyset} = 0$	
$K_{v,\{m_2\}} = 1$		$K_{u,\{m_1\}} = 2$	

$\eta(u)$	$\eta(v)$	$K_{u,\dots}$	$K_{v,\dots}$	valeur focale de u	valeur focale de v
0	0	$K_{u,\emptyset}$	$K_{v,\{m_1\}}$	0	1
0	1	$K_{u,\{m_2\}}$	$K_{v,\{m_1\}}$	2	1
1	0	$K_{u,\emptyset}$	$K_{v,\emptyset}$	0	0
1	1	$K_{u,\{m_2\}}$	$K_{v,\emptyset}$	2	0
2	0	$K_{u,\emptyset}$	$K_{v,\emptyset}$	0	0
2	1	$K_{u,\emptyset}$	$K_{v,\emptyset}$	0	0

FIGURE 1.5 – Tableaux de ressources et des points focaux pour un jeu de paramètres compatibles avec le graphe d'interaction présenté dans la figure 1.2. En haut, les valeurs choisies de paramètres. En bas, la liste des 6 états discrets, les paramètres correspondant aux ressources et la valeur des points focaux.

Avec les informations de la figure 1.5, nous pouvons construire le graphe d'états de la figure 1.6. On remarque que dans les états (0,1), (2,0) et (2,1) la distance entre la valeur courante de la variable u et de son point focal est de 2, et que par conséquent, les transitions horizontales partant de ces deux états ont été tronquées pour ne passer qu'un seul seuil. De même, on remarque que dans les états (1,1) et (2,1) les deux variables ont des points focaux différents de l'état courant. Par désynchronisation, il y a donc 2 transitions partant de ces états.

L'analyse de la figure 1.6 permet de dégager en particulier deux comportements remarquables : un cycle de longueur 4 (flèches rouges) et un autre la longueur 2 (flèches bleues).

- Sur la figure 1.2, on trouve un cycle entre u et v dans lesquels l'action de u sur v porte une négation alors que celle de v sur u n'en porte pas. Il y a donc un nombre impair de négations dans ce cycle et donc un nombre impair d'inhibitions. Un tel cycle est dit négatif. On constate également que l'atome indiquant l'action de u sur v change de valeur de vérité entre les seuils 0 et 1 et il en est de même pour celui qui indique l'action de v sur u . L'intersection entre ces deux seuils est appelée l'état caractéristique du cycle négatif. Le cycle de transition en rouge est centré autour de cet état caractéristique. C'est en fait le cycle d'interaction entre u et v qui le

génère.

- De la même façon, l'auto-régulation de u sur lui-même passe par une négation, cette fois au seuil 2. Lorsque $v = 0$, le multiplexe est faux quelle que soit la valeur de u . En revanche, lorsque $v = 1$, l'auto-inhibition de u est fonctionnelle. Puisque la valeur de u « décide » de la véracité du multiplexe. C'est en fait cette auto-inhibition qui génère le cycle bleu.

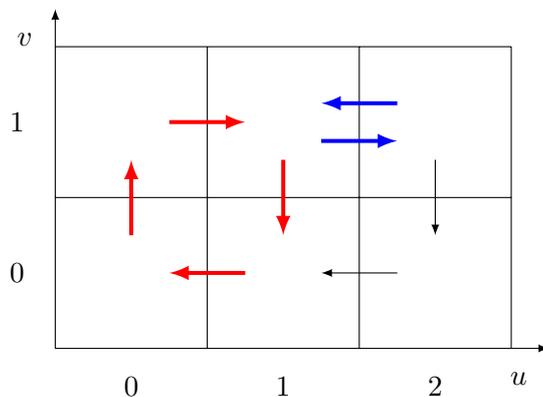


FIGURE 1.6 – Dynamique du réseau de la figure 1.2 avec les paramètres de la figure 1.5.

Houssine Snoussi a proposé des contraintes qui, quand elles sont satisfaites, assurent la compatibilité de la dynamique d'un réseau de Thomas avec le comportement des équations différentielles linéaires par morceaux [64]. Ces contraintes imposent que l'ajout d'une ressource ne fasse pas diminuer la valeur du paramètre correspondant. Autrement dit, pour $v \in V$, ω sous-ensemble de prédécesseurs de v et $n \in [0 \dots b_v]$, si $K_{v,\omega} = n$, alors $K_{v,\omega \cup \{m\}} \geq n$ avec m prédécesseur de v tel que $m \notin \omega$. Dans beaucoup de cas, ces contraintes sont cohérentes biologiquement puisque l'ajout d'un activateur ne fait généralement pas diminuer la vitesse de synthèse.

En ce point de la section, on a simplement exposé rapidement le formalisme de Thomas sous un angle compatible avec la logique formelle informatique. Ce formalisme discret est très simple du fait de son haut niveau d'abstraction. Il comporte également un relativement faible nombre de paramètres, en particulier grâce à l'utilisation des multiplexes qui permettent des factorisations efficaces. On va voir maintenant que grâce à son aspect formel, il est possible d'effectuer une identification automatique des paramètres par ordinateur en fonction des propriétés biologiques connues.

Identification des paramètres

Pour que le modèle ait un comportement biologiquement réaliste, il est nécessaire de prendre en compte toutes les connaissances biologiques et de sélectionner tous les modèles satisfaisant toutes ces propriétés. Pour automatiser cette identification de paramètres, une plateforme, *SMBioNet*, a été développée [44]. La figure 1.7 représente le fonctionnement de ce programme. Pour effectuer cette identification, il a besoin d'un graphe d'interac-

tion (1) et de propriétés issues des connaissances biologiques exprimées en logique temporelle CTL [15] (2). Il y a alors un prétraitement (flèche entre 2 et 3) qui a pour but de réduire le nombre de combinaisons de paramètres : les états stables, les boucles fonctionnelles sont des propriétés qui permettent de fixer les paramètres. Par exemple, pour les états stables, nous savons que la valeur de chacune des composantes est égale à la valeur de son paramètre. De plus, le programme autorise d'imposer les contraintes de Snoussi, ce qui permet de ne pas traiter les combinaisons de paramètres ne les satisfaisant pas. Une fois le prétraitement terminé, SMBioNet énumère l'ensemble des combinaisons de paramètres du réseau (3), construit les graphes d'états et vérifie si les modèles satisfont la conjonction des propriétés CTL grâce au model checker NuSMV [20] (4). Le programme permet donc d'obtenir l'ensemble exhaustif des modèles compatibles avec les connaissances biologiques (5).

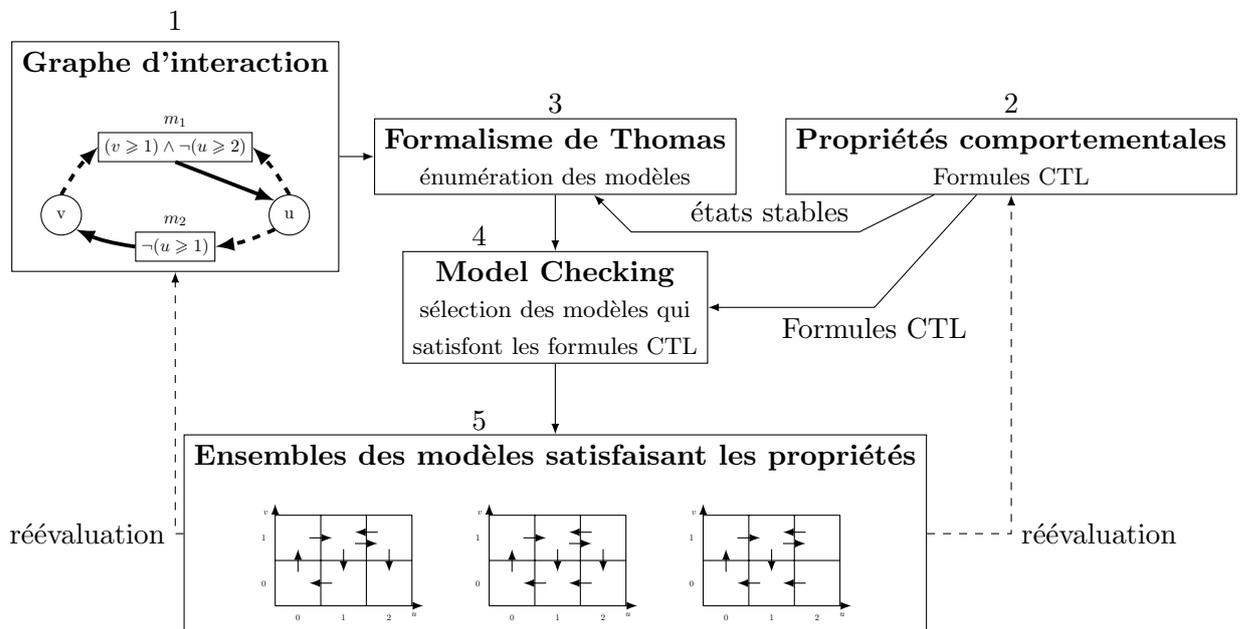


FIGURE 1.7 – Fonctionnement de la plateforme SMBioNet

Ce type d'identifications de paramètres n'est pas possible dans le formalisme des équations différentielles, car, d'une part ce n'est pas un formalisme symbolique, et d'autre part, le nombre de valeurs possibles pour un paramètre est infini.

1.2 Contributions principales de la thèse

Comme déjà mentionné, dans cette thèse, nous souhaitons développer des approches qui offrent la possibilité de concevoir des modèles de petite taille, fortement explicatifs mais suffisamment précis pour conserver un pouvoir de prédictibilité intéressant pour la biologie. Dans cette optique, cette thèse comporte essentiellement deux volets.

— Le premier volet développe une approche formelle pour réduire la taille d'un réseau

de Thomas discret sans en changer les comportements limites : le chapitre 2 définit l’approche, le chapitre 3, très technique, développe une preuve structurée de préservation des comportements limites. Le lecteur pourra, dans une première lecture, survoler ce chapitre. Enfin le chapitre 4 applique ce formalisme au réseau responsable du switch génétique chez le phage lambda.

- Le second volet définit un nouveau formalisme qui étend le formalisme discret de Thomas en prenant en compte une notion de temps continu. Ceci permet de raisonner formellement sur des délais avec un temps *chronométrique* (par opposition au temps *chronologique* de l’approche purement discrète) : le chapitre 5 fournit les définitions formelles de notre cadre de modélisation, le chapitre 6 développe l’exemple du cycle circadien et étudie les premières propriétés remarquables de sa dynamique, le chapitre 7 présente le logiciel `CircaPlot`, développé durant cette thèse qui permet d’affiner l’identification des paramètres, enfin le formalisme que nous avons développé dans ce second volet a déjà été repris par d’autres chercheurs et le chapitre 8 offre un survol de ces contributions.

Les deux sous-sections suivantes fournissent quelques détails supplémentaires sur ces deux volets.

1.2.1 Réduction formelle de réseaux de régulation génétiques

La plupart des réseaux de régulation génétique qui contrôlent un mécanisme biologique donné mettent en jeu un très grand nombre de gènes et donc de variables. Lorsque l’on conçoit un réseau génétique dédié à l’étude d’une fonction biologique, on est donc amené, dans un premier temps, à construire un réseau de taille très importante. L’identification des valeurs des paramètres compatibles avec les connaissances biologiques est alors impossible à effectuer dans un temps raisonnable : le nombre de combinaisons à énumérer est incommensurablement trop grand pour des réseaux dépassant une vingtaine ou une trentaine de variables. Il est peu probable que l’augmentation des puissances de calcul et les avancées sur les formalismes de modélisation permettent de dépasser des réseaux ayant quelques dizaines de variables. Il est donc nécessaire de travailler sur des réseaux plus petits. La solution qui semble s’imposer est donc d’augmenter encore le niveau d’abstraction pour construire un nouveau réseau plus petit qui reste pertinent compte tenu des questions abordées. Mettre en place un cadre permettant de prouver qu’un réseau plus abstrait répond de la même façon qu’un grand réseau à une question de n’importe quelle nature n’est pas chose aisée. Nous allons donc nous restreindre à des questions portant sur les comportements limites du système.

La méthode présentée dans cette thèse est une méthode de réduction formelle de réseaux qui permet de supprimer des informations non importantes vis-à-vis des comportements limites. Elle est inspirée des travaux d’Aurélien Naldi [54] qui a été le premier à développer une méthode de suppression de variables dans un réseau de Thomas (et

à notre connaissance, le seul). La méthode proposée par Naldi est essentiellement une méthode qui travaille au niveau du graphe d'états et des paramètres K du formalisme de Thomas. Nous allons exploiter la version formelle du formalisme de Thomas, exposée précédemment, pour exprimer les idées d'Aurélien Naldi sous une forme plus facilement exploitable. à cette occasion, nous allons étendre l'approche et résoudre en grande partie une limitation importante de la méthode de Naldi, à savoir le fait qu'il n'est pas possible de supprimer une variable autorégulée. Cette limitation était plus importante qu'il n'y paraît parce que les suppressions successives de variables créent de nombreuses boucles d'autorégulation.

Nous avons donc posé un cadre formel à la méthode de Naldi en utilisant les multiplexes, ce qui la rend plus simple, puis nous avons développé une extension qui permet, d'une part de supprimer des variables, mais également de supprimer les seuils sur lesquels il n'y a pas d'autorégulation. Nous avons donc repoussé la limitation de la méthode de Naldi, car il nous est possible de réduire une variable autorégulée sans pour autant la supprimer. Tout comme la méthode de Naldi, notre méthode permet de conserver les comportements asymptotiques (états stables et attracteurs cycliques). Même si se limiter aux comportements asymptotiques est relativement restrictif, ces premiers résultats d'abstractions de réseaux de Thomas ont de l'intérêt, car les comportements asymptotiques sont facilement identifiables par l'expérimentation. De plus, dans le cadre des modèles discrets, ils constituent des informations particulièrement importantes.

Afin d'illustrer notre formalisme, nous l'avons appliqué au réseau génétique responsable du switch génétique du phage λ . Nous avons montré le lien entre deux modèles construits de manière indépendante à partir d'observations biologiques. L'un comporte 4 variables et l'autre 2. La réduction formelle du réseau à 4 variables nous a permis de retrouver celui à 2 variables, montrant ainsi qu'il en est une abstraction valide vis-à-vis des comportements limites.

1.2.2 Formalisme de modélisation hybride

Ce formalisme a été motivé par une collaboration avec l'équipe de Franck Delaunay du laboratoire IBV (Institut de Biologie de Valrose - Université de Nice Sophia Antipolis) dont le domaine de recherche est le cycle circadien chez les mammifères. Les premiers travaux effectués en début de thèse ont par ailleurs contribué à l'obtention du contrat ANR « Hyclock » qui regroupe également l'équipe INSERM *Rythmes Biologiques et Cancers* à Villejuif, l'équipe CNRS de l'école centrale de Nantes *MeForBio* et l'équipe *BIOCHAM* de Inria Rocquencourt.

De manière naturelle, la modélisation du cycle circadien requiert de savoir combien de temps le système reste dans un état discret donné. Du fait que l'approche de Thomas ne permet d'étudier que la succession des changements d'états discrets, il nous fallait y ajouter

une notion de temps *chronométrique*. Les formalismes existants imposaient des paramètres de délais en plus des paramètres discrets de l'approche de Thomas. Cela engendrait une forte lourdeur dans l'étape d'identification des paramètres. L'idée qui motive la définition de ce nouveau formalisme est la suivante : *la connaissance des vitesses de changements de taux de concentration permet de déduire la valeur des paramètres discrets de Thomas*. Nous avons alors développé un formalisme plus simple dans lequel ces vitesses remplacent les paramètres de Thomas, ce qui nous a permis de conserver la simplicité de l'approche de Thomas.

Pour démontrer à la fois la faisabilité et la puissance de ce nouveau formalisme, nous avons développé un modèle de l'horloge circadienne qui respecte l'essentiel des propriétés biologiques connues (résistance aux décalages horaires, adaptation aux variations saisonnières ...). De manière assez remarquable, seulement deux variables ont été nécessaires pour atteindre ce résultat (sans compter les entrées lumineuses). Ceci montre que l'ajout de nos vitesses à la théorie de Thomas constitue un outil fiable et précis pour reproduire des comportements réalistes avec des modèles très simples.

Au-delà du cycle circadien, de nombreux systèmes physiologiques sont soumis à des cycles temporels (souvent pilotés par le noyau suprachiasmatique) de sorte que notre approche est utile pour formaliser bon nombre d'autres réseaux avec un temps chronométrique. Ces modèles sont importants pour la chronopharmacologie¹ et, dans le cadre de l'ANR Hyclock par exemple, un modèle du cycle cellulaire a également été défini en utilisant notre formalisme hybride. Enfin, toujours dans le cadre de Hyclock, une extension du formalisme de la « logique de Hoare génétiquement modifiée » [16] a été définie pour notre formalisme hybride, ce qui renforce l'intérêt d'une telle définition formelle avec vitesse.

1. qui étudie l'impact du choix du moment de la prise d'un médicament

Chapitre 2

Réduction de réseaux discrets de Thomas

Lorsqu'on modélise un réseau génétique grâce à la théorie de R. Thomas, les paramètres ne sont généralement pas connus et leur identification par énumération est trop longue de par le nombre exponentiel de paramétrisations à considérer. Cette étape d'identification est cruciale et un temps de calcul trop important constitue un frein pour l'étude de la dynamique des réseaux génétiques.

Pour pallier cette limitation, le modélisateur doit alors proposer des graphes d'interactions plus petits préservant les comportements qui l'intéressent. L'expertise du système biologique permet de délimiter un sous-ensemble de gènes importants pour l'hypothèse posée. Cependant, la taille du réseau reste souvent trop élevée pour pouvoir identifier les paramètres à l'aide d'un ordinateur dans un temps de calcul raisonnable. Un des objectifs de la réduction de réseau génétique consiste donc à *abstraire* certains gènes, comme un gène dont le seul rôle est de servir de relais, et ce, de manière assistée par ordinateur. Cette méthodologie a d'abord été introduite par Aurélien Naldi dans le cadre des réseaux de Thomas [53, 54].

Ce chapitre a pour objectif de définir et formaliser une extension de la méthode d'Aurélien Naldi. Après avoir brièvement présenté sa méthode, nous montrerons que les opérations de réduction peuvent être codées dans les formules des multiplexes : les formules des multiplexes du réseau réduit sont formellement déduites du réseau initial. Cette extension est fondée sur les idées d'Aurélien Naldi et est applicable dans le cas où l'on cherche à fusionner deux niveaux consécutifs d'une variable.

2.1 Approche de réduction de Naldi

2.1.1 Intuition

La méthode de réduction de Naldi se place dans le cadre de modélisation des réseaux de Thomas. L'intuition est simple : supprimer une variable revient à raccorder les prédécesseurs du gène supprimé à ses successeurs. La Figure 2.1, représente l'effet de la suppression de la variable r sur le graphe d'interaction : malheureusement, le nombre de paramètres peut largement augmenter. Dans la figure de gauche (avant la réduction), il y a 24 paramètres ($16 = 2^4$ pour r et $8 = 2^3$ pour v) alors que dans la figure de droite (après réduction), il y a $64 = 2^6$ paramètres.

L'intuition sous-jacente de cette méthode est la suivante : on fait comme si la variable à supprimer évoluait très vite par rapport aux autres variables. Ainsi, à partir d'un état η , elle passe rapidement de sa valeur courante ($\eta(v)$) à sa valeur focale ($K_{v,\rho(\eta,v)}$) avant que les autres variables n'aient eu le temps d'évoluer.

La Figure 2.2 illustre l'effet de la réduction sur la dynamique d'un réseau (à gauche la dynamique d'origine, et à droite la dynamique après la suppression de la variable r). Les

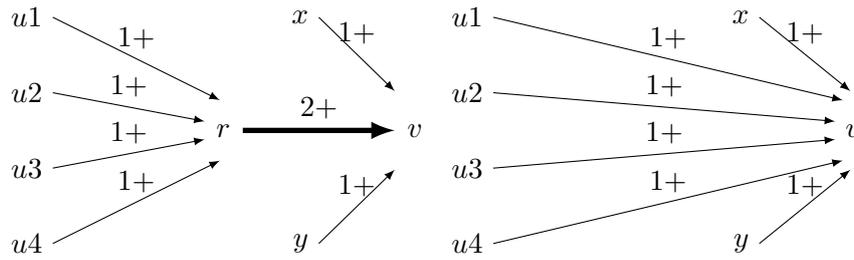


FIGURE 2.1 – Intuition de la méthode de réduction de Naldi. La notation $x \xrightarrow{s+} y$ représente le multiplexe $\boxed{x \geq s} \rightarrow y$

flèches horizontales dans la dynamique initiale correspondent aux transitions très rapides de la variable que l'on supprime. Lorsqu'on entre dans une boîte de couleur, on ne peut en sortir que lorsqu'on ne peut plus emprunter de transitions horizontales.

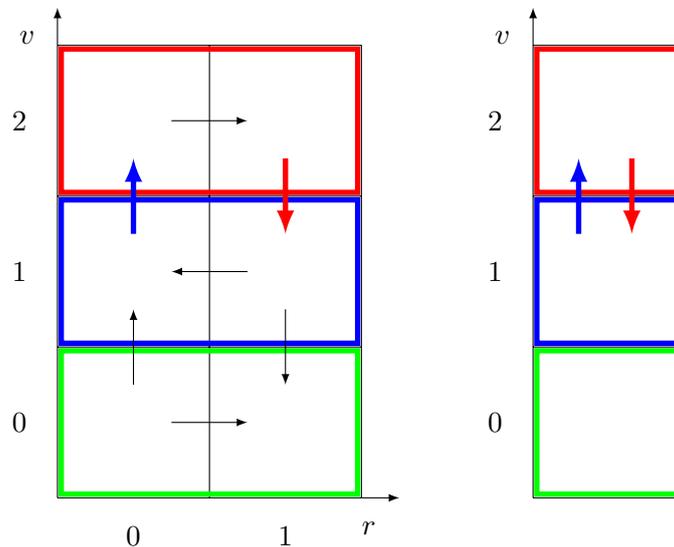


FIGURE 2.2 – Effet de la suppression de la variable r sur la dynamique d'un réseau. À gauche, la dynamique du réseau non réduit, à droite la dynamique du réseau réduit.

Naturellement, à l'intérieur d'une boîte, il faut que la valeur focale pour v de chacun des états soit identique, car, dans le cas contraire, on ne saurait pas vers quelle valeur aller. Le paramètre $K_{r,\dots}$ (la valeur focale de r) doit donc être le même dans tous les états d'une boîte. Ce paramètre dépend de l'ensemble des ressources qui peut lui-même contenir un multiplexe dont le littéral contient r lorsqu'il y a autorégulation.

La Figure 2.3 illustre la dynamique autour d'autorégulation négative (gauche) et positive (droite). Dans le cas d'une autorégulation négative, le graphe d'états contiendra un cycle de longueur 2 autour du seuil (en dessous du seuil, la variable augmente, et au-dessus, elle diminue), et dans le cas d'une autorégulation positive, aucune transition ne traversera le seuil (en dessous du seuil, la variable ne peut pas augmenter, alors qu'au-dessus elle ne peut pas diminuer).

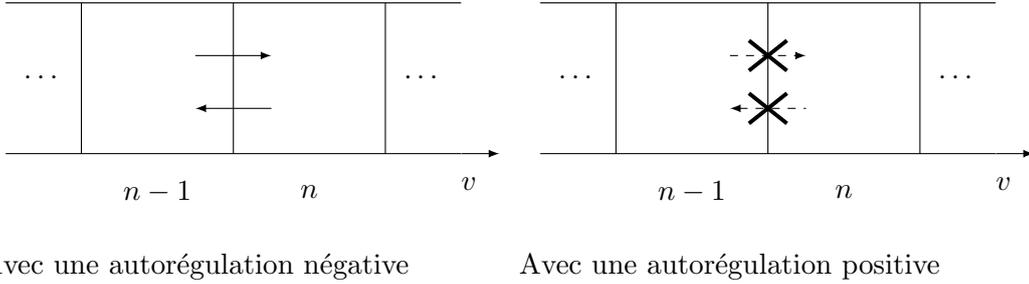


FIGURE 2.3 – Dynamique locale autour d'un seuil d'autorégulation.

Pour être plus précis, lorsque la variable est autorégulée à un seuil s donné (il existe un multiplexe qui régule r dont un des atomes de la formule est de la forme $r \geq s$), les ressources de r ne seront pas les mêmes selon la valeur de $\eta(r)$ par rapport à s . Il n'y a pas unicité de la valeur focale lorsque l'on fait varier la valeur de r . On ne peut donc pas choisir la valeur focale vers laquelle $r \ll$ avance très vite \gg .

- Dans le cas d'une variable booléenne et d'une autorégulation négative (Figure 2.4 bas à gauche), la valeur focale est atteignable. Cependant, une fois la valeur focale atteinte, l'état correspondant ayant lui même une autre valeur focale, il faudrait aller rapidement vers cette nouvelle valeur focale, et on se retrouve au point de départ. Ainsi, la multiplicité des valeurs focales empêche de savoir vers quelle valeur ira le système. Le cas d'une boucle positive est d'autant plus problématique (Figure 2.4 bas à droite), car la variable r est déjà à sa valeur focale et le problème de la multiplicité des valeurs focales est le même.
- Dans le cas d'une variable multivaluée (Figure 2.4 haut et milieu), les valeurs focales ne sont pas forcément atteignables. Aller directement vers ces valeurs focales permettrait, dans le réseau réduit, un saut qualitatif vers ces valeurs, ce qui ne correspondrait à aucun chemin dans le réseau non réduit.

Plus concrètement, lorsqu'une variable r est auto-activée au seuil s (la formule du multiplexe m est évaluée à vrai si $r \geq s$), le passage de la valeur $s - 1$ à la valeur s (pour r alors que les autres variables ne changent pas) ajoute la ressource m à l'ensemble ω des ressources de r . Ce changement d'ensemble de ressources entraîne un changement du paramètre ($K_{r,\omega} \neq K_{r,\omega'}$ avec $\omega' = \omega \cup \{m\}$). Ces deux paramètres n'ont aucune raison d'avoir la même valeur. Le cas d'une autorégulation négative s'explique de manière similaire.

Ainsi, nous ne pouvons supprimer une variable ou fusionner des états seulement lorsque la valeur focale à considérer n'est pas ambiguë. Autrement dit, l'unicité de la valeur focale de la variable à supprimer est indispensable à l'automatisation des réductions.

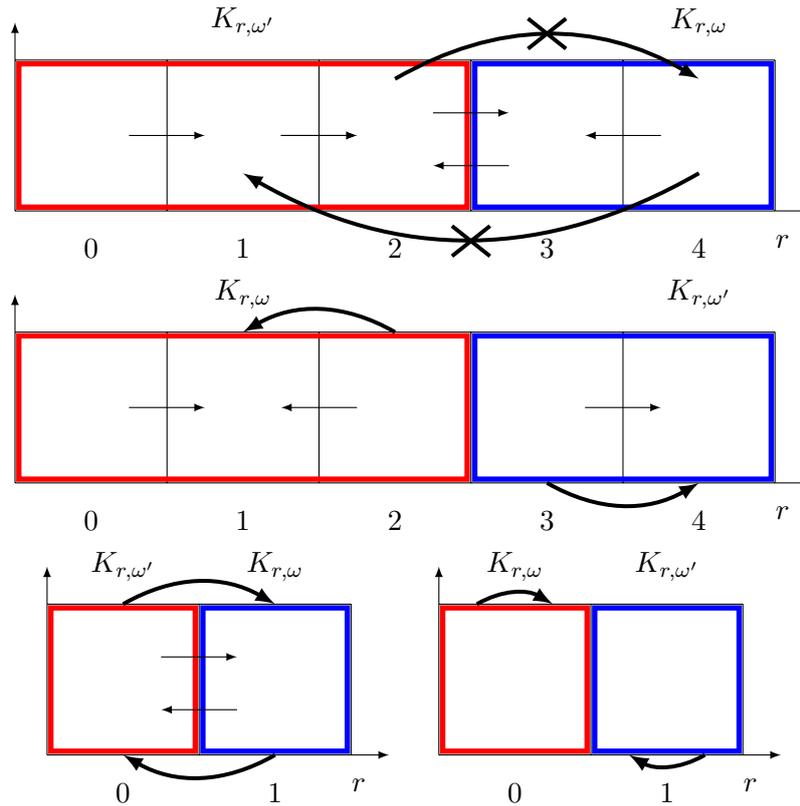


FIGURE 2.4 – Comportement dans le cas d’une autorégulation. En haut : autorégulation négative. Au milieu : autorégulation positive. En bas, dans le cas d’une variable booléenne : à gauche, autorégulation négative et à droite, positive. Tous les états d’une même boîte de couleur ont les mêmes ressources pour r (le cadre rouge a pour ressource ω et le cadre bleu, ω'), ils ont donc les mêmes valeurs focales pour r . Les flèches épaisses représentent les sauts qualitatifs vers ces valeurs focales qui, quand elles sont barrées, ne correspondent à aucun chemin de proche en proche.

2.1.2 Reformulation de la méthode de Naldi

Dans cette section, nous donnons une définition constructive du réseau réduit à la manière de Naldi. Cette définition repose sur des formules logiques, alors que dans l’article d’origine, Aurélien Naldi présente son approche du point de vue algorithmique.

Notons que nos définitions permettent de construire le réseau réduit à la manière de Naldi dans le cadre de modélisation de réseaux de régulation avec multiplexes. Ainsi, pour reformuler les idées de Naldi (qui n’avaient pas de multiplexes), nous restreignons le langage des multiplexes à un seul atome de type $v \geq n$ muni éventuellement d’une négation qui représentera le signe $-$ classiquement utilisé par Thomas. Lorsque plusieurs prédécesseurs contribueront à l’activation d’une cible unique, les conditions de coopération seront oubliées, menant à un nombre de paramètres beaucoup plus grand. Nous dirons alors qu’une variable r est autorégulée s’il existe un multiplexe m dont la formule est réduite à l’atome $r \geq s$ ou à sa négation.

Savoir si une variable tend vers une valeur est une propriété qu'on doit vérifier de manière récurrente dans la construction du réseau réduit. Cette propriété est représentée grâce à la formule donnée dans la définition suivante.

Définition 6. Soit $A = (V, M, E, \mathcal{K})$ un réseau de régulation génétique. Considérons une variable $v \in V$, un entier n tel que $0 \leq n \leq b_v$, et un état η . La formule $F_{v,n}(\eta)$ est définie par :

$$F_{v,n}(\eta) = \bigvee_{\omega \text{ t.q. } K_{v,\omega}=n} \rho(\eta, v) = \omega \quad \text{où } \omega \text{ est un sous-ensemble des prédécesseurs de } v$$

Cette formule est évaluée à vrai si à un état donné η , la variable v tend vers la valeur n . Au maximum, il ne peut y avoir qu'un seul élément de la disjonction qui soit vrai, car, à l'état η , l'ensemble de ressources $\rho(\eta, v)$ est unique. Dans le cas où le paramètre $K_{v,\rho(\eta,v)}$ n'est pas égal à n , la disjonction sera vide et la formule fausse.

Notation : Soient une variable v , un état η et un entier $n \in [0 \dots b_v]$. La notation $\eta[v \leftarrow n]$ représente l'état η' dans lequel $\eta'(v) = n$ et pour toute $u \neq v$, $\eta'(u) = \eta(u)$.

La définition suivante fait intervenir la notion de littéral qui représente soit un atome, soit sa négation. Les formules des multiplexes sont donc réduites à des littéraux.

Définition 7. Soit $A = (V, M, E, \mathcal{K})$ un réseau dont les formules de tous les multiplexes sont réduites à des littéraux. Soit une variable $v \in V$ et une autre variable $r \neq v$ non autorégulée. Soit n un entier tel que $0 \leq n \leq b_v$. Soit η un état du réseau A et η^r sa projection sur $V \setminus \{r\}$.

Soit P l'ensemble des multiplexes dont le littéral contient r et qui ont v comme successeur.

La formule $F_{v,n}^r(\eta)$ suivante satisfaite pour l'état η , si à l'état η^r (qui est la projection de η sur $V \setminus \{r\}$) du réseau réduit, les ressources de v lui permettent de tendre vers n . Elle est définie par :

$$F_{v,n}^r(\eta) = \begin{cases} F_{v,n}(\eta) & \text{si } P = \emptyset \\ \bigvee_{\omega \text{ t.q. } K_{v,\omega}=n} \left(\bigvee_{x \text{ t.q. } \rho(\eta[r \leftarrow x], v) = \omega} F_{r,x}(\eta) \right) & \text{sinon} \end{cases}$$

Si la disjonction est vide, la formule est fausse.

L'intuition est la suivante. Pour que $F_{v,n}^r(\eta)$ soit vraie, il faut que le point focal de r à l'état η^r permette à v de tendre vers n .

- Dans le cas où r n'a aucune action sur v , c'est la même formule qu'avant la réduction (lorsque $P = \emptyset$, la définition donne $F_{v,n}(\eta)$).
- Dans le cas où r a une action sur v , il faut que, si r atteint sa valeur focale très rapidement, les ressources dans l'état où r a glissé jusqu'à sa valeur focale, permettent à v de tendre vers n ($\rho(\eta[r \leftarrow x], v) = \omega$ et $K_{v,\omega} = n$).

Remarquons que dans la formule $F_{v,n}^r(\eta)$, dans le cas où P est non vide, il y a au plus un seul terme de la disjonction qui est vrai : comme on l'a vu précédemment, en η , la variable r tend vers une valeur x ($F_{r,x}(\eta)$), et en $\eta[r \leftarrow x]$, l'ensemble des ressources $\rho(\eta[r \leftarrow x], v)$ est unique. S'il n'existe pas de x tel que $K_{v,\rho(\eta[r \leftarrow x],v)} = n$, la disjonction est vide. Comme l'ensemble des ressources de v en $\eta[r \leftarrow x]$ est unique et que les paramètres $K_{v,\dots}$ sont définis pour tout ensemble de ressources, il n'existe qu'un n tel que $F_{v,n}^r(\eta^r)$ soit satisfaite.

Nous sommes maintenant en mesure de définir la valeur des paramètres du réseau réduit.

Définition 8. Soit $A = (V, M, E, \mathcal{K})$ un réseau dont les formules de tous les multiplexes sont réduites à des littéraux. Soient deux variables v et r de V différentes et r non autoréglée. Soit η un état du réseau A et η^r sa projection sur $V \setminus \{r\}$. La valeur focale de la variable v dans le réseau réduit vaut $K_{v,\rho(\eta^r,v)}^r = n$ si $F_{v,n}^r(\eta)$ est satisfaite.

Notation 1. Pour plus de clarté, nous utiliserons la notation $K_v^r(\eta^r)$ à la place de $K_{v,\rho(\eta^r,v)}^r$

Comme dit précédemment, il existe un unique n tel que $F_{v,n}^r(\eta^r)$ soit évaluée à vrai. Par conséquent, $K_v^r(\eta^r)$ est bien défini. On peut maintenant construire le réseau réduit obtenu par suppression d'une variable r non autoréglée.

Définition 9. Soit $A = (V^A, M^A, E^A, \mathcal{K}^A)$ un RRG tel que tous les multiplexes contiennent un unique littéral de la forme $x \geq n$ ou $\neg(x \geq n)$. Soit $r \in V^A$ une variable non autoréglée, autrement dit, telle qu'aucun multiplexe qui pointe sur r n'ait une formule qui contient un atome de la forme $r \geq k$.

Le réseau B obtenu par suppression de la variable r est défini par :

$$\begin{aligned} & - V^B = V^A \setminus \{r\} \\ & - M^B = \left\{ m_{u,v}^k \left| \begin{array}{l} u \in V^B, v \in V^B, 1 \leq k \leq b_u \\ \exists \eta^r \text{ tel que } \eta^r(u) = k - 1, \exists n \in [0 \dots b_v] \text{ tel que } F_{v,n}^r(\eta^r) \neq F_{v,n}^r(\eta^r + \epsilon^u) \\ \text{où } \epsilon^u \text{ est le vecteur de base dont la seule composante non nulle est } u. \end{array} \right. \right\} \end{aligned}$$

La formule associée à $m_{u,v}^k$ est

$$\begin{aligned} & \left\{ \begin{array}{ll} (u \geq k) & \text{si } K_v^r(\eta^r) < K_v^r(\eta^r + \epsilon^u) \text{ ou} \\ \neg(u \geq k) & \text{si } K_v^r(\eta^r) > K_v^r(\eta^r + \epsilon^u) \end{array} \right. \\ & - E^B = \{m_{u,v}^k \rightarrow v \mid m_{u,v}^k \in M^B\} \\ & - K^B = \{K_{v,\omega}^B\} \text{ est la famille des entiers, indexée par} \\ & \quad - v \in V^B \text{ et} \\ & \quad - \omega \subset \{m_{u,v}^k \in M^B \mid u \in V, k \in [1, \dots, b_u]\} \end{aligned}$$

Les entiers $K_{v,\omega}^B$ sont définis par :

$$K_{v,\omega}^B = k \quad \text{si} \quad (\rho^B(\eta^r, v) = \omega) \Rightarrow F_{v,k}^r(\eta^r)$$

La définition précédente fait intervenir ρ^B qui est bien défini, car le calcul des ressources ne dépend pas des paramètres K_{\dots}^B .

La suppression d'une variable r sur un réseau modifie chacun des éléments du quadruplet.

- L'ensemble des variables V^B ne contient naturellement plus la variable r tout en conservant les autres.
- L'ensemble des multiplexes M^B contient chacun des multiplexes $m_{u,v}^k$ qui représentent l'action de u sur v au niveau k . Si u passe de $k - 1$ à k , la valeur focale de v change. La formule du multiplexe sera $u \geq k$ si la valeur focale augmente, et $\neg(u \geq v)$ si elle diminue.
- Chaque multiplexe $m_{u,v}^k$ pointe vers son successeur v , ce qui forme l'ensemble des arcs E^B .
- Le paramètre $K_{v,\omega}^B$ prendra la valeur k si pour tous les états où l'ensemble des ressources de v est ω , v est attiré vers k dans le réseau réduit ($F_{v,k}^r(\eta)$). L'article d'Aurélien Naldi ne donne pas de preuve de l'unicité de la valeur focale pour un ensemble de ressources donné. Dans la suite de ce chapitre, nous proposons une définition de la réduction qui contourne cette difficulté.

La méthode de réduction d'Aurélien Naldi préserve les comportements asymptotiques [54], à savoir les points fixes, les attracteurs cycliques et les attracteurs complexes composés de plusieurs cycles entremêlés.

Notons par ailleurs que l'espace des états est lui aussi largement réduit : la suppression de la variable r amène à un réseau dont le nombre d'états a été divisé par $b_r + 1$.

2.1.3 Les limites de la méthode

La principale limitation de cette méthode de réduction est l'explosion du nombre de paramètres à chaque suppression de variable (dans la figure 2.1, le nombre de paramètres passe de 24 à 64). En effet, une réduction peut générer un grand nombre de multiplexes pointant vers une même variable. S'il y a un multiplexe $m_{u,v}^k$ qui représente l'action de u sur v au seuil k , on peut aussi avoir des multiplexes $m_{u,v}^l$ avec $l \neq k$.

De plus, à chacune des réductions, tous les paramètres sont redéfinis, même si certains multiplexes se retrouvent inchangés. La phase de reconstruction de ces multiplexes qui alourdit la définition 9 devrait pouvoir être contournée au prix éventuel d'une modification des formules des multiplexes.

Enfin, un réseau dans lequel toutes les variables sont autorégulées ne peut pas être réduit. Il peut être qualifié de *réseau minimal* puisque cette méthode ne peut pas le réduire d'avantage. Notons qu'un enchaînement de réductions est susceptible de générer des autorégulations. Par exemple, la suppression d'une variable d'un cycle de longueur 3

va donner un cycle de longueur 2, qui à son tour, peut donner lieu à un cycle de longueur 1, c'est-à-dire à une autorégulation. Du point de vue de la dynamique, un cycle fonctionnel de longueur 3 génère un comportement cyclique de longueur 8. Le cycle de longueur 2 obtenu aura un comportement cyclique de longueur 4 et l'autorégulation aura un comportement cyclique de longueur 2.

Les réductions successives vont mener à un réseau minimal dans lequel toutes les variables sont autorégulées. Cependant, des séquences de réductions différentes ne donneront pas nécessairement le même réseau minimal [31].

La suite de ce chapitre met à profit la puissance des formules des multiplexes pour coder les réductions par suppression de gènes non autorégulés. Les opérations nécessaires à ces réductions vont être codées par des formules intégrées aux multiplexes. De plus, nous étendons cette notion de réduction au cas des variables autorégulées à un seuil donné : bien évidemment, on ne pourra pas supprimer ces variables mais diminuer leurs ensembles de valeurs possibles.

2.2 Construction et formalisation du réseau réduit

Rappelons que la méthode de réduction de Naldi rend impossible la suppression d'une variable autorégulée du fait de l'existence de plusieurs valeurs focales existantes. Cependant, tous les états ayant les mêmes ressources (c'est-à-dire tous les états situés entre deux seuils successifs sur lesquels se situent les autorégulations) ont une unique valeur focale. Dans la Figure 2.5, il y a une autorégulation sur le seuil 2, donc il y a 2 valeurs focales : les états 0, 1 et 2 d'un côté ont la même valeur focale et les états 3 et 4 de l'autre ont aussi la même valeur focale. Le point rouge pour les états 0,1,2 et le point bleu pour les états 3, 4 représentent les valeurs focales associées à ces deux ensembles d'états. Intuitivement, les états de chacun de ces deux ensembles peuvent être fusionnés, puisque dans chacun d'eux, la variable r peut se diriger, sans sauts qualitatifs non justifiés, en direction de sa valeur focale (jusqu'à éventuellement, sortir de l'ensemble d'états considéré).

Plus précisément, deux états adjacents qui ont les mêmes ressources ont la même valeur focale et peuvent donc être fusionnés en un état ayant pour valeur focale de la variable, celle qui est commune aux deux états initiaux. La méthode de réduction définie dans cette section est basée sur cette opération de suppression de seuil ne portant pas d'autorégulation.

Avant de définir la réduction de réseau par suppression de seuil, nous définissons des formules qui vont faciliter l'écriture des multiplexes du réseau réduit. Ces formules expriment la sémantique des idées fondatrices de R. Thomas.

Définition 10. Soient $A = (V, M, E, \mathcal{K})$ un RRG, $v \in V$, $s \in [1, \dots, b_v]$ un seuil de v et

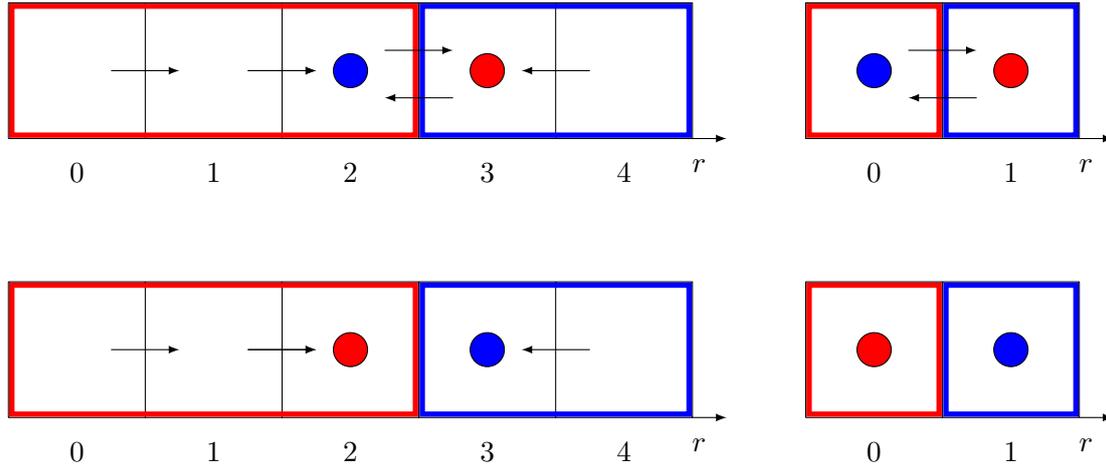


FIGURE 2.5 – Comportements dans les ensembles d'états ayant les mêmes ressources dans le cas d'une autorégulation négative (en haut) et positive (en bas). Pour les deux types d'autorégulations, à gauche, les dynamiques d'origine et à droite les dynamiques après réduction des sous-ensembles d'états

ω un ensemble de prédécesseurs de v . Nous définissons les 3 formules suivantes :

a. $\Phi_s^v \equiv (v \geq s - 1) \wedge \neg(v \geq s + 1)$

Cette formule est évaluée à vrai si la valeur de v dans l'état courant est égale à $s - 1$ ou s . Sémantiquement, elle est vraie pour les états adjacents au seuil s et est équivalente à $(v = s) \vee (v = s - 1)$.

b. $\Phi_\omega^v \equiv (\bigwedge_{m \in \omega} \varphi_m) \wedge (\bigwedge_{m \in A^{-1}(v) \setminus \omega} \neg \varphi_m)$

Cette formule est évaluée à vrai si ω est l'ensemble des ressources de v à l'état courant.

c. $\Phi^{v \geq s} \equiv \bigwedge_{\omega \subset A^{-1}(v)} (\Phi_\omega^v \implies K_{v, \omega} \geq s)$

Cette formule est évaluée à vrai si les ressources de v à l'état courant lui permettent d'être attirée vers une valeur au-dessus de s .

La suppression d'un seuil amène à faire quelques changements dans la numérotation des états, et par conséquent, dans les paramètres et les formules. Lorsque l'on supprime le seuil s de la variable v , si $\eta(v) < s$, la composante v de η après réduction restera identique, mais si $\eta(v) \geq s$, la composante v de l'état réduit sera égale à $\eta(v) - 1$. Il en est de même pour les paramètres $K_{v, \dots}$ et pour les termes v présents dans les formules des multiplexes. Ainsi, quand une valeur liée à une variable v ($\eta(v)$, $K_{v, \dots}$ ou v dans une formule de multiplexe) est supérieure au seuil supprimé, elle doit diminuer de 1.

D'une manière générale, lorsqu'on traverse un seuil de la variable v ne portant pas d'autorégulation, les ressources de v ne changent pas. C'est ce qu'exprime le lemme suivant.

Lemme 1. Soit $A = (V^A, M^A, E^A, \mathcal{K}^A)$ un RRG, $r \in V^A$ une variable non autorégulée au seuil $s \in [1 \dots b_v]$. Deux états adjacents tels que $\eta(v) = s$ et $\eta'(v) = s - 1$ ont les mêmes

ressources.

$$\forall \eta \in S^A, \rho(\eta[r \leftarrow s - 1], r) = \rho(\eta[r \leftarrow s], r)$$

Puisqu'il n'y a pas d'autorégulation, les ressources ne peuvent pas changer après le passage du seuil. La preuve est donc évidente.

Lors d'une réduction, la suppression d'un seuil d'une variable v réduit le nombre de valeurs possibles pour v de 1 ($b_v - 1$). Il faut donc replier le réseau de façon à ce qu'aucune valeur v , $\eta(v)$ et $K_{v,\dots}$ ne corresponde à une valeur qui n'existe pas.

Définition 11. [Fonctions de repliement de réseau] Soient $A = (V^A, M^A, E^A, \mathcal{K}^A)$ un RRG, $r \in V^A$ une variable et $s \in [1, \dots, b_r]$ un seuil sur lequel il n'y a pas d'autorégulation sur r et $t \in \mathbb{N}$. Soit $\eta^A \in S^A$ un état de A . Enfin, considérons l'ensemble des états du réseau réduit $B : S^B = [0, b_r - 1] \times \prod_{x \in V \setminus \{r\}} [0, b_x]$

Les fonctions de repliement sont définies par :

$$\text{— } fold_s^r : S^A \rightarrow S^B \text{ définie par } fold_s^r(\eta^A) = \begin{cases} \eta^A[\eta(r) \leftarrow \eta(r) - 1] & \text{si } \eta^A(r) \geq s \\ \eta^A & \text{sinon} \end{cases}$$

$$\text{— } fold_k_s^r : \mathbb{N} \rightarrow \mathbb{N} \text{ définie par } fold_k_s^r(K_{a,\omega}) = \begin{cases} K_{a,\omega} - 1 & \text{si } a = r \text{ et } K_{a,\omega} \geq s \\ K_{a,\omega} & \text{sinon} \end{cases}$$

— $fold_f_s^r : \mathcal{F} \rightarrow \mathcal{F}$, où \mathcal{F} est l'ensemble des formules bien formées, définie inductivement par :

$$\begin{aligned} & \bullet \text{ Si } \varphi \text{ est de la forme } \begin{cases} r \geq t \text{ avec } t \geq s, \text{ alors } fold_f_s^r(\varphi) = r \geq t - 1 \\ K_{r,\omega} \geq t \text{ avec } t \geq s : \\ \text{— Si } b_r \geq 2, \text{ alors } fold_f_s^r(\varphi) = fold_k_s^r(K_{r,\omega}) \geq t - 1 \\ \text{— Si } b_r < 2, \text{ alors } fold_f_s^r(\varphi) = \top \end{cases} \\ & \bullet \text{ Si } \varphi \text{ est un autre atome, alors } fold_f_s^r(\varphi) = \varphi \\ & \bullet \text{ Si } \varphi \text{ est de la forme } \begin{cases} \neg\psi, \text{ alors } fold_f_s^r(\varphi) = \neg fold_f_s^r(\psi) \\ \psi_1 \wedge \psi_2, \text{ alors } fold_f_s^r(\varphi) = fold_f_s^r(\psi_1) \wedge fold_f_s^r(\psi_2) \\ \psi_1 \vee \psi_2, \text{ alors } fold_f_s^r(\varphi) = fold_f_s^r(\psi_1) \vee fold_f_s^r(\psi_2) \\ \psi_1 \Rightarrow \psi_2, \text{ alors } fold_f_s^r(\varphi) = fold_f_s^r(\psi_1) \Rightarrow fold_f_s^r(\psi_2) \end{cases} \end{aligned}$$

Pour simplifier, nous noterons ces 3 fonctions avec le même nom : $fold_s^r$.

D'après la définition précédente, lors de la suppression du seuil d'une variable booléenne r , le repliement d'un atome de la forme $K_{r,\omega} \geq n$ est la formule \top . En effet, supposons que l'on fasse le même repliement que dans le cas où $b_r \geq 2$. La formule obtenue serait $fold_k_s^r(K_{r,\omega}) \geq 1 - 1 = 0$. Par conséquent, après le repliement de l'atome, la formule sera toujours satisfaite.

Définition 12. [Réduction de réseaux de régulations génétiques] Soit $A = (V^A, M^A, E^A, \mathcal{K}^A)$ un RRG, r une variable et $s \in [1, \dots, b_r]$ un seuil de r sur lequel

$$fold_s^r \left((\neg\Phi_s^r \wedge \varphi_m^A) \vee (\Phi_s^r \wedge \varphi_m^A[r \geq s \leftarrow \Phi^{r \geq s}]) \right)$$

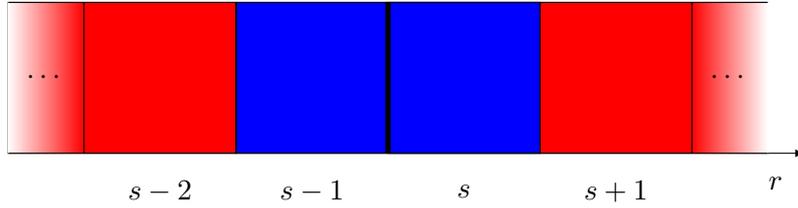


FIGURE 2.6 – Correspondance entre les parties de la nouvelle formule des multiplexes et les états de la variable traitée, dans le cas multivalué.

il n'y a pas d'autorégulation. Le RRG obtenu par suppression du seuil s de r est le réseau $B = (V^B, M^B, E^B, \mathcal{K}^B)$ défini par :

1. Si $b_r^A = 1$:
 - $V^B = V^A \setminus \{r\}$
 - $M^B = M^A \setminus \{m \mid m \rightarrow r \in E^A, \forall v \neq r, m \rightarrow v \notin E^A\}$ et $\forall m \in M^B, \varphi_m^B \equiv fold_s^r(\varphi_m^A[r \geq s \leftarrow \Phi^{r \geq s}])$
 - $E^B = E^A \setminus \{m \rightarrow r \mid m \in M^A\}$
 - $\mathcal{K}^B = \mathcal{K}^A \setminus \{K_{r,\omega}^A \mid \omega \subset A^{-1}(r)\}$
2. Si $b_r^A \geq 2$:
 - $V^B = V^A$, $b_r^B = b_r^A - 1$ et $\forall v \in V^A \setminus \{r\}$, $b_v^B = b_v^A$
 - $M^B = M^A$, et $\forall m \in M^B$, $\varphi_m^B \equiv fold_s^r((\neg\Phi_s^r \wedge \varphi_m^A) \vee (\Phi_s^r \wedge \varphi_m^A[r \geq s \leftarrow \Phi^{r \geq s}]))$
 - $E^B = E^A$
 - $\mathcal{K}^B = \mathcal{K}^A \setminus \{K_{r,\omega}^A \mid \omega \subset A^{-1}(r)\} \cup \{K_{r,\omega}^B = fold_s^r(K_{r,\omega}^A) \mid \omega \subset B^{-1}(r)\}$

Dans la définition du réseau réduit, deux cas sont à distinguer :

- Si la variable simplifiée r est booléenne (suppression du seuil $s = 1$), elle est supprimée de l'ensemble des variables ainsi que ses paramètres et ses multiplexes prédécesseurs s'ils n'ont que r comme successeur. De plus, pour chacun des multiplexes restants, chacun des atomes de la forme $r \geq s$ est remplacé par la formule $\Phi^{r \geq s}$ (vraie si r peut passer son seuil à l'état courant). Dans ce cas particulier, supprimer un seuil est équivalent à supprimer la variable.
- Si la variable simplifiée r est multivaluée, la borne de r est réduite et les anciens paramètres de r sont remplacés par les paramètres repliés. Chacune des formules sera aussi modifiée. Dans le cas où on est loin du seuil supprimé ($\neg\Phi_s^r$, en rouge dans la Figure 2.6), la formule ne change pas, et si l'on est adjacent au seuil (Φ_s^r), les atomes de la forme $r \geq s$ sont remplacés par $\Phi^{r \geq s}$ (en bleu dans la Figure 2.6).

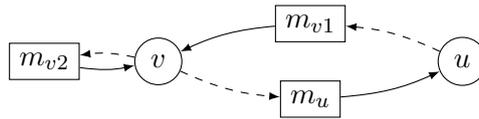
Dans le cas où une variable n'est pas autorégulée, tous les seuils de cette va-

riable peuvent être supprimés successivement. Les opérations de suppression de variable d'Aurélien Naldi sont donc applicables avec le formalisme présenté ici.

De la même façon que dans le formalisme de Naldi, la réduction présentée ici nécessite de connaître les paramètres. Cependant, l'adaptation du processus de réduction à un nouveau jeu de paramètres est facilitée par l'évaluation des formules modifiées des multiplexes. Il n'est plus nécessaire de reconstruire les multiplexes au prix d'une complexité des formules importante.

2.3 Exemple de réduction

Afin d'illustrer ce formalisme de réduction, nous allons l'appliquer sur un réseau à 2 variables : une multivaluée (3 valeurs) et l'autre booléenne. La Figure 2.7 contient la description du réseau.



$V^A :$	$M^A :$	$E^A :$	$\mathcal{K}^A :$	
$v (b_v = 2)$	$m_{v1} : \neg(u \geq 1)$	$(m_{v1} \rightarrow v)$	$K_v = 0$	$K_u = 0$
$u (b_u = 1)$	$m_{v2} : \neg(v \geq 2)$	$(m_{v2} \rightarrow v)$	$K_{v,\{m_{v1}\}} = 1$	$K_{u,\{m_u\}} = 1$
	$m_u : \neg(v \geq 1)$	$(m_u \rightarrow u)$	$K_{v,\{m_{v2}\}} = 0$	
			$K_{v,\{m_{v1},m_{v2}\}} = 2$	

FIGURE 2.7 – Réseau génétique à 2 variables.

La Figure 2.8 contient le tableau des valeurs focales du réseau de la figure 2.7 (en haut) ainsi qu'une représentation de la dynamique. Cette dynamique contient deux comportements asymptotiques : un point fixe à l'état $v = 0$ et $u = 1$ et un cycle de longueur 2 entre les états $(v = 1, u = 0)$ et $(v = 2, u = 0)$.

Dans ce réseau, nous ne pouvons supprimer que deux seuils : le seuil 1 de v et l'unique seuil de u . Comme u est une variable booléenne, la suppression de son seuil supprimera la variable u . À l'inverse, le seuil 2 de v contient une autorégulation et ne peut pas être supprimé.

2.3.1 Suppression du seuil 1 de u

La variable u étant booléenne, supprimer son seuil 1 revient à supprimer la variable. Il faut donc traiter tous les multiplexes contenant des occurrences de u , autrement dit, le

$\eta(v)$	$\eta(u)$	Paramètres de v	Paramètres de u	Valeurs focales
0	0	$K_{v,\{m_{v1},m_{v2}\}}$	$K_{u,\{m_u\}}$	2 1
0	1	$K_{v,\{m_{v2}\}}$	$K_{u,\{m_u\}}$	0 1
1	0	$K_{v,\{m_{v1},m_{v2}\}}$	$K_{u,\emptyset}$	2 0
1	1	$K_{v,\{m_{v2}\}}$	$K_{u,\emptyset}$	0 0
2	0	$K_{v,\{m_{v1}\}}$	$K_{u,\emptyset}$	1 0
2	1	$K_{v,\emptyset}$	$K_{u,\emptyset}$	0 0

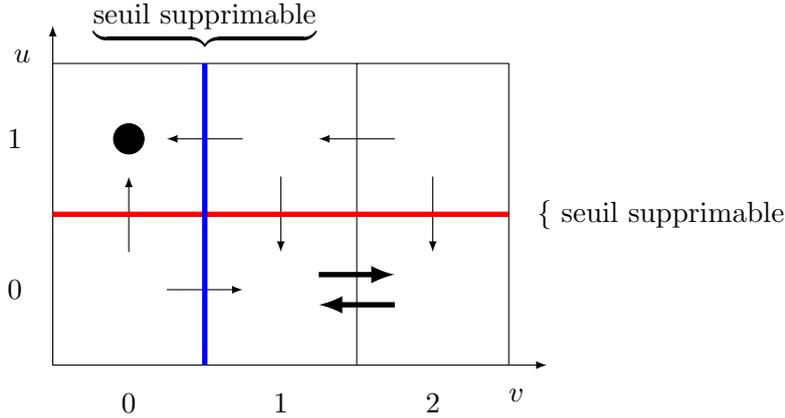


FIGURE 2.8 – Dynamique du réseau représenté dans la Figure 2.7.

multiplexe m_{v1} . D'après la définition 12, $\varphi_{m_{v1}}$ devient $fold_u^1(\varphi_{m_{v1}}[u \geq 1 \leftarrow \Phi^{u \geq 1}])$.

$$\begin{aligned}
 fold_u^1(\varphi_{m_{v1}}[u \geq 1 \leftarrow \Phi^{u \geq 1}]) &\equiv fold_u^1(\neg \Phi^{u \geq 1}) \\
 &\equiv fold_u^1(\underbrace{\neg((\Phi_{\emptyset}^v \implies K_{v,\emptyset} \geq 1))}_{\alpha} \wedge \underbrace{(\Phi_{\{m_u\}}^v \implies K_{v,\{m_u\}} \geq 1)}_{\beta})
 \end{aligned}$$

La sous-formule β est vraie, car $K_{v,\emptyset} \geq 1$ est vraie. La formule précédente devient :

$$\begin{aligned}
 fold_u^1(\neg(\varphi_{m_{v1}}[u \geq 1 \leftarrow \Phi^{u \geq 1}])) &\equiv fold_u^1(\neg(\Phi_{\emptyset}^v \implies K_{v,\emptyset} \geq 1)) \\
 &\equiv fold_u^1(\neg((v \geq 1) \implies K_{v,\emptyset} \geq 1)) \quad \text{car } \Phi_{\emptyset}^v \equiv \neg(\neg(v \geq 1)) \\
 &\equiv fold_u^1(\neg(\neg(v \geq 1) \vee \underbrace{K_{v,\emptyset} \geq 1}_{\perp})) \\
 &\equiv fold_u^1(\neg(\neg(v \geq 1))) \equiv v \geq 1
 \end{aligned}$$

La nouvelle formule de m_{v1} devient après simplification ($v \geq 1$). Le réseau réduit est décrit dans les figures 2.9 et 2.10.

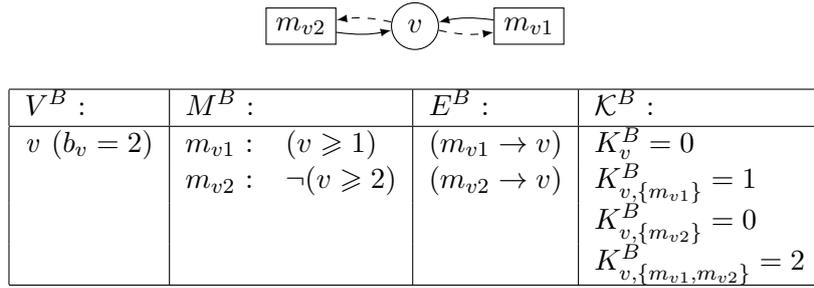
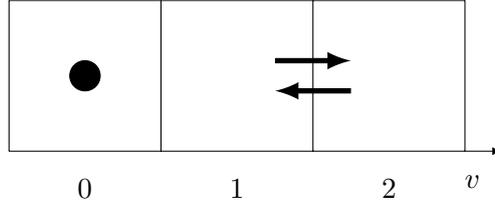
FIGURE 2.9 – Réseau génétique de la figure 2.7 obtenu par suppression de la variable u 

FIGURE 2.10 – Dynamique du réseau représenté dans la Figure 2.9.

2.3.2 Suppression du seuil 1 de v

La variable v ne peut pas être supprimée en une seule étape pour deux raisons : elle est multivaluée et elle est autorégulée sur le seuil 2. Cependant, la méthode de réduction peut être appliquée sur le seuil 1 de v . Nous calculons le réseau réduit $C = (V^C, M^C, E^C, \mathcal{K}^C)$ obtenu par suppression du seuil 1 de v . Les 3 ensembles V^C , E^C et \mathcal{K}^C se déduisent facilement :

- L'ensemble des variables ne change pas : $V^C = V^A$. Par contre la borne b_v passe de 2 à 1.
- L'ensemble des arcs ne change pas : $E^C = E^A$.
- Les paramètres $K_{v,\omega}^A$ sont repliés. Or les valeurs de ces paramètres dans le réseau initial valent : $K_v^A = 0$, $K_{v,\{m_{v1}\}}^A = 1$, $K_{v,\{m_{v2}\}}^A = 0$ et $K_{v,\{m_{v1},m_{v2}\}}^A = 2$. En appliquant les définitions 11 et 12, ces paramètres deviennent $K_v^C = 0$, $K_{v,\{m_{v1}\}}^C = 0$, $K_{v,\{m_{v2}\}}^C = 0$ et $K_{v,\{m_{v1},m_{v2}\}}^C = 1$.

Toutes les formules des multiplexes vont être repliées, mais seul le multiplexe m_u sera drastiquement affecté. En effet, il est le seul à avoir un atome de la forme $v \geq 1$. Cet atome sera remplacé par la formule $\Phi^{v \geq 1}$. Dans le réseau réduit C , φ_{m_u} devient :

$$\varphi_{m_u}^C \equiv \text{fold}_1^v(\underbrace{(\neg\Phi_1^v \wedge \varphi_{m_u})}_{\alpha}) \vee (\Phi_1^v \wedge \varphi_{m_u}[v \geq 1 \leftarrow \Phi^{v \geq 1}])$$

La sous-formule α est évaluée à faux. En effet, d'après la définition 10, $\neg\Phi_1^v$ est équivalent à $\neg(v \geq 0) \vee (v \geq 2)$. Comme $\neg(v \geq 0)$ est faux, $\neg\Phi_1^v$ est équivalente à $(v \geq 2)$ et Φ_1^v est équivalente à $\neg(v \geq 2)$. Ainsi, on a $\neg\Phi_1^v \wedge \varphi_{m_u} \equiv (v \geq 2) \wedge \neg(v \geq 1) \equiv \perp$. La

formule $\varphi_{m_u}^C$ devient :

$$\varphi_{m_u}^C \equiv fold_1^v(\neg(v \geq 2) \wedge \varphi_{m_u}[v \geq 1 \leftarrow \Phi^{v \geq 1}])$$

Développons la formule $\Phi^{v \geq 1}$:

$$\Phi^{v \geq 1} \equiv \begin{cases} \Phi_{\emptyset}^v & \Rightarrow K_v \geq 1 & (\alpha) \\ \wedge \Phi_{\{m_{v1}\}}^v & \Rightarrow K_{v,\{m_{v1}\}} \geq 1 & (\beta) \\ \wedge \Phi_{\{m_{v2}\}}^v & \Rightarrow K_{v,\{m_{v2}\}} \geq 1 & (\gamma) \\ \wedge \Phi_{\{m_{v1},m_{v2}\}}^v & \Rightarrow K_{v,\{m_{v1},m_{v2}\}} \geq 1 & (\epsilon) \end{cases}$$

Puisque $a \Rightarrow b$ est équivalent à $\neg(a) \vee b$, les implications (β) et (ϵ) sont toujours vraies, car d'après les paramètres du réseau donné dans la figure 2.7, $K_{v,\{m_{v1}\}} = 1$ et $K_{v,\{m_{v1},m_{v2}\}} = 2$. A l'inverse les termes contenant les paramètres $K_{v,\dots}$ dans (α) et (γ) sont faux (cf. valeurs des paramètres de la figure 2.7). On a donc :

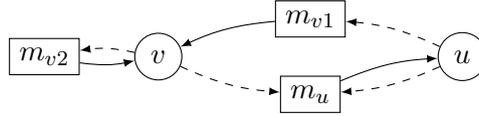
$$\begin{aligned} \Phi^{v \geq 1} &\equiv \neg\Phi_{\{\emptyset\}}^v \wedge \neg\Phi_{\{m_{v2}\}}^v \\ &\equiv \underbrace{\neg((v \geq 2) \wedge (u \geq 1))}_{\Phi_{\{\emptyset\}}^v} \wedge \underbrace{\neg(\neg(v \geq 2) \wedge (u \geq 1))}_{\Phi_{\{m_{v2}\}}^v} \\ &\equiv (\neg(v \geq 2) \vee \neg(u \geq 1)) \wedge ((v \geq 2) \vee \neg(u \geq 1)) \\ &\equiv \neg((u \geq 1) \vee (\neg(v \geq 2) \wedge (v \geq 2))) \\ &\equiv \neg(u \geq 1) \end{aligned}$$

La formule $\varphi_{m_u}^C$ devient :

$$\begin{aligned} \varphi_{m_u}^C &\equiv fold_1^v(\neg(v \geq 2) \wedge \neg(\neg(u \geq 1))) \\ &\equiv fold_1^v(\neg(v \geq 2) \wedge (u \geq 1)) \\ &\equiv \neg(v \geq 1) \wedge (u \geq 1) \end{aligned}$$

Le réseau réduit obtenu est décrit dans les figures 2.11 et 2.12.

Nous pouvons constater, au travers de ces deux illustrations de la méthode de réduction de seuil, qu'à chaque fois, le réseau réduit (R^B ou R^C) possède un point fixe et un attracteur



V^C :	M^C :	E^C :	\mathcal{K}^C :
v ($b_v = 1$)	$m_{v1} : \neg(u \geq 1)$	$(m_{v1} \rightarrow v)$	$K_v^C = 0$ $K_u^C = 0$
u ($b_u = 1$)	$m_{v2} : \neg(v \geq 1)$	$(m_{v2} \rightarrow v)$	$K_{v,\{m_{v1}\}}^C = 0$ $K_{u,\{m_u\}}^C = 1$
	$m_u : \neg(v \geq 1) \wedge (u \geq 1)$	$(m_u \rightarrow u)$	$K_{v,\{m_{v2}\}}^C = 0$
			$K_{v,\{m_{v1},m_{v2}\}}^C = 1$

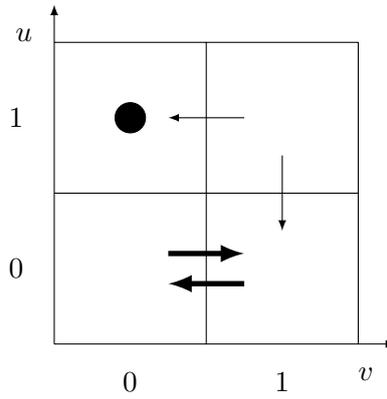
FIGURE 2.11 – Réseau réduit de la figure 2.7 par suppression du seuil 1 de v 

FIGURE 2.12 – Dynamique du réseau représenté dans la Figure 2.11.

cyclique de taille 2 comme le réseau initial.

2.4 Discussion et conclusion

Dans ce chapitre, nous avons pu, grâce aux multiplexes, formaliser la réduction présentée par Aurélien Naldi. De plus, alors que sa méthode génère un grand nombre de paramètres différents pour conserver les comportements d'intérêt, notre méthode contient dans les multiplexes, les informations sur les abstractions faites lors de la réduction. Le nombre de paramètres *diminue* donc au fur et à mesure de la simplification du modèle.

Notre méthode permet également de supprimer un seuil d'une variable, ce qui permet de dépasser les limitations de la méthode de Naldi : il est possible de diminuer le nombre d'états même si toutes les variables sont autorégulées.

Cependant, cette méthode n'est pas symbolique. Du fait que les paramètres sont des entiers connus, les réductions ne sont faisables que si tous les paramètres sont identifiés. Si les paramètres deviennent symboliques, nous pourrions effectuer les réductions qui seraient valables, quelque soit le jeu de paramètres au prix d'une perte drastique des simplifications des formules des multiplexes. Les bases de cette méthode symbolique sont en cours de

réduction.

Une seconde limitation réside dans l'impossibilité de retrouver le réseau d'origine à partir du réseau réduit. L'utilisation d'une méthode symbolique de réduction donnerait la possibilité d'effectuer cette « remontée ». Durant les réductions, le lien entre le réseau réduit et celui d'origine pourrait être mémorisé, ce qui permettrait de retrouver le réseau d'origine à partir du réseau réduit. Par conséquent, nous pourrions identifier les paramètres sur le réseau réduit, puis identifier des contraintes sur les paramètres du réseau initial afin de retrouver des comportements similaires sur le réseau d'origine. Nous verrons dans le chapitre suivant que les comportements conservés sont les trajectoires asymptotiques.

Chapitre 3

Préservation des propriétés par réduction

Pour la biologie, les comportements asymptotiques constituent une information très importante. En effet, lors d'une expérimentation, ces comportements sont détectables même avec une faible précision temporelle (relevés d'observation espacés). Par exemple, la dynamique du réseau responsable de la morphogenèse de *Arabidopsis thaliana* est composée principalement d'états stables correspondant à des tissus différents (sépalés, pétales, carpelles et étamines) résultant de processus de différenciation menant à la formation de la fleur [51, 26].

Dans ce chapitre, nous montrerons que les réductions de réseau par suppression de seuils préservent les propriétés dynamiques asymptotiques. Ainsi, ce type de réduction peut permettre au biologiste de raisonner sur un réseau plus petit dans lequel il est certain que les comportements asymptotiques facilement observables sont présents.

Ce chapitre est exclusivement technique. Il donne les preuves des différents lemmes utiles pour les théorèmes de conservation des états stables et des attracteurs cycliques lors d'une réduction. La figure 3 représente les liens de dépendance entre les lemmes et théorèmes : les lemmes 2 à 7 sont nécessaires aux théorèmes 1 (conservation des états stables) et 2 (conservation des attracteurs cycliques).

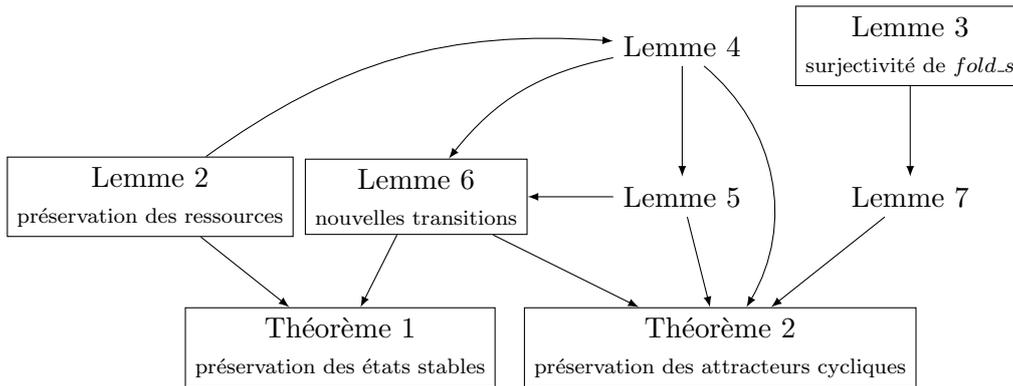


FIGURE 3.1 – Schéma général des preuves des théorèmes 1 et 2

Le lemme 2 exprime le fait que les ressources d'un état non adjacent au seuil qui sera supprimé sont conservées lors de la réduction.

Lemme 2. [Préservation des ressources] Soit $A = (V^A, M^A, E^A, \mathcal{K}^A)$ un RRG, $G^A = (S^A, T^A)$ son graphe d'états, $\eta^A \in S^A$ un état et s un seuil de la variable v sans autorégulation. Soit $B = (V^B, M^B, E^B, \mathcal{K}^B)$ le réseau réduit obtenu par suppression du seuil s de v , et $G^B = (S^B, T^B)$ son graphe d'états.

Si $\eta^A \not\models \Phi_s^v$, alors pour chaque variable $u \in V$, nous avons $\rho^A(\eta^A, u) = \rho^B(\text{fold}_s^v(\eta^A), u)$

Démonstration. Soient φ_m^A et φ_m^B les formules du multiplexe m dans les réseaux A et B . Soit un état $\eta^B \in S^B$ tel que $\eta^B = \text{fold}_s^v(\eta^A)$. Comme $\eta^A \not\models \Phi_s^v$, d'après la définition des multiplexes du réseau réduit (Définition 12) $\varphi_m^B \equiv \text{fold}_s^v(\neg\Phi_s^v \wedge \varphi_m^A) \equiv \text{fold}_s^v(\varphi_m^A)$. Donc, nous avons $\eta^B \models \varphi_m^B \Leftrightarrow \eta^B \models \text{fold}_s^v(\varphi_m^A)$.

Prouvons que $\eta^A \models \varphi_m^A \Leftrightarrow \eta^B \models \text{fold}_s^v(\varphi_m^A)$:

1. Cas où φ_m^A ne contient aucun atome contenant v .

Puisque φ_m^A ne dépend pas de v , la réduction ne change pas l'évaluation de cette formule (pas de substitution). Par conséquent, $\text{fold}_s^v(\varphi_m^A) \Leftrightarrow \varphi_m^A$.

De plus, pour $u \neq v$, la valeur de la variable u ne change pas entre η^A et η^B , donc $\eta^B(u) = \eta^A(u)$. Par conséquent, les atomes ne changent pas. Nous avons bien $\eta^A \models \varphi_m^A \Leftrightarrow \eta^B \models \text{fold}_s^v(\varphi_m^A)$.

2. Si φ_m^A contient au moins un atome contenant la variable v .

— Si $\eta^A(v) < s$, pour tout $u \in V^A$, $\eta^B(u) = \text{fold}_s^v(\eta^A(u)) = \eta^A(u)$. En utilisant le fait que $\eta^A(v) < s$, $\eta^B \models \text{fold}_s^v(\varphi_m^A) \Leftrightarrow \eta^B \models \varphi_m^A$, nous avons $\eta^A \models \varphi_m^A \Leftrightarrow \eta^B \models \text{fold}_s^v(\varphi_m^A)$.

— Si $\eta^A(v) \geq s$, nous avons $\eta^B = \eta^A[v \leftarrow v - 1]$.

— Le repliement des atomes de la forme $v \geq t$ avec $t < s$ n'affecte pas leur évaluation. Donc le repliement de $\eta^A(v)$ n'affectera pas l'évaluation de ces atomes.

— Le repliement des atomes de la forme $v \geq t$ avec $t \geq s$ ne modifie pas leur évaluation. En effet,

— si $v < s$, l'atome ($v \geq t$) est replié en ($v \geq t + 1$) et l'évaluation de ces atomes est inchangée,

— si $v > s$, le repliement de l'atome $v \geq s$ donne $v - 1 \geq s - 1$. Ainsi l'évaluation de ces atomes est inchangée.

Donc, nous avons $\eta^A \models \varphi_m^A \Leftrightarrow \eta^B \models \text{fold}_s^v(\varphi_m^A) \Leftrightarrow \eta^B \models \varphi_m^B$

Si $\eta^B \models \varphi_m^B$, alors $m \in \rho^B(\eta^B, u)$ d'après la définition des ressources (Définition 4 du chapitre 1, page 18). Comme $\eta^A \models \varphi_m^A \Leftrightarrow \eta^B \models \varphi_m^B$, m est ressource de u à η^A si et seulement si m est ressource de u à η^B . Plus généralement, nous avons $\rho^A(\eta^A, u) = \rho^B(\text{fold}_s^v(\eta^A), u)$. \square

Le lemme 3 est un intermédiaire technique. Il exprime le fait que le repliement de l'espace des états est surjectif. En effet, les états adjacents au seuil supprimé sont « fusionnés » lors de la réduction.

Lemme 3. Soit $A = (V^A, M^A, E^A, \mathcal{K}^A)$ un RRG, $v \in V^A$ une variable de A et s un seuil de v (sur lequel il n'y a pas d'autorégulation). La fonction de repliement fold_s^v est surjective.

Démonstration. Soit $G^A = (S^A, T^A)$ le graphe d'états du réseau A . Soit B le réseau obtenu par réduction du seuil s de v et $G^B = (S^B, T^B)$ son graphe d'états. Soit η^B un état de S^B . Construisons un état $\eta^A \in S^A$ tel que $\text{fold}_s^v(\eta^A) = \eta^B$:

- Pour tout $u \neq v$, on pose $\eta^A(u) = \eta^B(u)$
- Pour la composante v de η^A ,
 - Si $\eta^B(v) < s - 1$, $\eta^A(v) = \eta^B(v)$
 - Si $\eta^B(v) > s - 1$, $\eta^A(v) = \eta^B(v) + 1$
 - Si $\eta^B(v) = s - 1$, on pose $\eta^A(v) = \eta^B(v)$ (on aurait pu aussi choisir $\eta^A(v) = \eta^B(v) + 1$). Donc nous avons $fold_s^v(\eta^A) = \eta^B$ et $\eta^A \in S^A$

□

Le lemme 4 exprime la préservation des transitions pour lesquels au moins un des états n'est pas adjacent au seuil supprimé. Ce lemme concerne toutes les transitions qui ne longent pas le seuil supprimé.

Lemme 4. Soient $A = (V^A, M^A, E^A, \mathcal{K}^A)$ un RRG, $v \in V^A$ une variable, s un seuil de v sur lequel il n'y a pas d'autorégulation sur v . Les transitions pour lesquelles au moins une des extrémités ne satisfait pas la propriété Φ_s^v sont préservées par la réduction du seuil s de v .

Démonstration. Soit $G^A = (S^A, T^A)$ le graphe d'états de A , $\eta^A, \eta'^A \in S^A$ deux états tels que $(\eta^A \rightarrow \eta'^A) \in T^A$, v et s la variable et le seuil de v de la réduction considérée. Il existe un unique u tel que $\eta^A(u) \neq \eta'^A(u)$ et $\eta'^A(u) = \eta^A(u) + \text{sgn}(K_{u,\rho^A(\eta^A,u)}^A - \eta^A(u))$ où $\text{sgn}(x) = 1$ si $x > 0$, $\text{sgn}(x) = 0$ si $x = 0$ et $\text{sgn}(x) = -1$ si $x < 0$.

Si $u \neq v$, comme $fold_s^v$ replie uniquement la variable v , u ne change pas. D'après le lemme 2, $\eta^B = fold_s^v(\eta^A)$ a les mêmes ressources que η^A , donc η^B est attiré dans la direction u vers $fold_s^v(K_{u,\rho(\eta^B,u)})$. Le repliement du paramètre $K_{u,\dots}$ ne modifie pas sa valeur, donc il y aura une transition dans le réseau réduit de η^B vers $fold_s^v(\eta'^A)$.

Dans le cas où la transition $\eta^A \rightarrow \eta'^A$ est sur la dimension de v , trois cas doivent être considérés.

1. Premier cas : $\eta^A \not\in \Phi_s^v$ et $\eta'^A \not\in \Phi_s^v$. D'après la définition du repliement, les images par $fold_s^v$ de deux états adjacents sont adjacentes. De plus,
 - Si $\eta^A(v) < s - 1$ et $\eta'^A(v) < s - 1$, $fold_s^v(\eta^A)(v) = \eta^A(v)$ et $fold_s^v(\eta'^A)(v) = \eta'^A(v)$.
 - Si $K_{v,\rho^A(\eta^A,v)}^A < s$, alors $K_{v,\rho^A(\eta^A,v)}^A = fold_s^v(K_{v,\rho^A(\eta^A,v)}^A)$, donc $fold_s^v(K_{v,\rho^A(\eta^A,v)}^A) < s$. La transition est conservée parce que les ressources sont préservées.
 - Si $K_{v,\rho^A(\eta^A,v)}^A \geq s$, alors $fold_s^v(K_{v,\rho^A(\eta^A,v)}^A) = K_{v,\rho^A(\eta^A,v)}^A - 1$, donc $fold_s^v(K_{v,\rho^A(\eta^A,v)}^A) \geq s - 1$. La transition est encore conservée (préservation des ressources).
 - Si $\eta^A(v) > s$ et $\eta'^A(v) > s$, $fold_s^v(\eta^A)(v) = \eta^A(v) - 1$ et $fold_s^v(\eta'^A)(v) = \eta'^A(v) - 1$.

- Si $K_{v,\rho^A(\eta^A,v)}^A < s$, alors $K_{v,\rho^A(\eta^A,v)}^A = fold_s^v(K_{v,\rho^A(\eta^A,v)}^A)$, alors $fold_s^v(K_{v,\rho^A(\eta^A,v)}^A) < s$. La transition est conservée (préservation des ressources).
- Si $K_{v,\rho^A(\eta^A,v)}^A \geq s$, alors $fold_s^v(K_{v,\rho^A(\eta^A,v)}^A) = K_{v,\rho^A(\eta^A,v)}^A - 1$, alors $fold_s^v(K_{v,\rho^A(\eta^A,v)}^A) \geq s - 1$. La transition est conservée (préservation des ressources).

2. Deuxième cas : $\eta^A \not\models \Phi_s^v$ et $\eta'^A \models \Phi_s^v$.

- (a) Supposons que $\eta'^A(v) = \eta^A(v) + 1$. Nous avons $\eta^A(v) = s - 2$ et $\eta'^A(v) = s - 1$. Alors $K_{v,\rho^A(\eta^A,v)}^A > s - 2$. Comme $\eta^A(v) < s$, $fold_s^v$ ne modifie pas $\eta^A(v)$.
- Si $K_{v,\rho^A(\eta^A,v)}^A = s - 1$, comme $K_{v,\rho^A(\eta^A,v)}^A$ n'est pas affecté par $fold_s^v$, $fold_s^v(K_{v,\rho^A(\eta^A,v)}^A) = s - 1$, la transition est conservée (préservation des ressources).
 - Si $K_{v,\rho^A(\eta^A,v)}^A \geq s$, nous avons $K_{v,\rho^A(\eta^A,v)}^A = fold_s^v(K_{v,\rho^A(\eta^A,v)}^A) + 1$ alors $fold_s^v(K_{v,\rho^A(\eta^A,v)}^A) + 1 > s - 2$. la transition est conservée (*idem*).
- (b) Supposons que $\eta'^A(v) = \eta^A(v) - 1$. Nous avons $\eta^A(v) = s + 1$ et $\eta'^A(v) = s$. Alors $K_{v,\rho^A(\eta^A,v)}^A < s + 1$. Comme $\eta^A(v) \geq s$, $fold_s^v$ ne modifie pas $\eta^A(v)$.
- Si $K_{v,\rho^A(\eta^A,v)}^A \leq s - 1$, comme $K_{v,\rho^A(\eta^A,v)}^A$ n'est pas affecté par $fold_s^v$, $fold_s^v(K_{v,\rho^A(\eta^A,v)}^A) \leq s - 1$, la transition est conservée.
 - Si $K_{v,\rho^A(\eta^A,v)}^A = s$, nous avons $K_{v,\rho^A(\eta^A,v)}^A = fold_s^v(K_{v,\rho^A(\eta^A,v)}^A) + 1$ alors $fold_s^v(K_{v,\rho^A(\eta^A,v)}^A) + 1 = s + 1$. la transition est conservée.

3. Troisième cas : $\eta^A \not\models \Phi_s^v$ et $\eta'^A \models \Phi_s^v$.

- (a) Supposons $\eta'^A(v) = \eta^A(v) + 1$. Nous avons $\eta^A(v) = s$, $\eta'^A(v) = s + 1$ et $K_{v,\rho^A(\eta^A,v)}^A > s$. Donc $K_{v,\rho^A(\eta^A,v)}^A = fold_s^v(K_{v,\rho^A(\eta^A,v)}^A) + 1$ et $fold_s^v(\eta^A)(v) = s - 1$.

Montrons maintenant que les ressources de v ne sont pas affectées par le repliement. En effet, soit m une ressource de v à l'état η^A . Le repliement de φ_m^A est $fold_s^v((\neg\Phi_s^v \wedge \varphi_m^A) \vee (\Phi_s^v \wedge \varphi_m^A[v \geq s \leftarrow \Phi^{v \geq s}]))$ qui est équivalent à $fold_s^v(\varphi_m^A[v \geq s \leftarrow \Phi^{v \geq s}])$. Ici, $v \geq s$ est vrai et $\Phi^{v \geq s}$ est vrai car v augmente. m est donc aussi une ressource de v à l'état $fold_s^v(\eta^A)$.

Puisqu'il y a préservation des ressources et que le repliement ne change pas les positions relatives de η^A , η'^A et $K_{v,\dots}$, nous avons $fold_s^v(\eta^A) \rightarrow fold_s^v(\eta'^A)$.

- (b) Supposons $\eta'^A(v) = \eta^A(v) - 1$. Nous avons $\eta^A(v) = s - 1$, $\eta'^A(v) = s - 2$ et $K_{v,\rho^A(\eta^A,v)}^A < s - 1$. Donc $K_{v,\rho^A(\eta^A,v)}^A = fold_s^v(K_{v,\rho^A(\eta^A,v)}^A)$ et $fold_s^v(\eta^A)(v) = s - 1$.

Montrons maintenant que les ressources de v ne sont pas affectées par le repliement. En effet, soit m une ressource de v à l'état η^A . Le repliement de φ_m^A est $fold_s^v((\neg\Phi_s^v \wedge \varphi_m^A) \vee (\Phi_s^v \wedge \varphi_m^A[v \geq s \leftarrow \Phi^{v \geq s}]))$ qui est équivalent à $fold_s^v(\varphi_m^A[v \geq s \leftarrow \Phi^{v \geq s}])$. Ici, $v \geq s$ est faux et $\Phi^{v \geq s}$ est faux, car v diminue. m est donc aussi une ressource de v à l'état $fold_s^v(\eta^A)$.

Puisqu'il y a préservation des ressources et que le repliement ne change pas les positions relatives de η^A , η'^A et $K_{v,\dots}$, nous avons $fold_s^v(\eta^A) \rightarrow fold_s^v(\eta'^A)$.

□

Le lemme 5 exprime le fait que, si à un état donné v n'évolue plus, toutes les transitions partant de cet état sont conservées.

Lemme 5. *Soit $G^A = (S^A, T^A)$ et $G^B = (S^B, T^B)$ les graphes d'états du réseau initial A et du réseau réduit B obtenu par suppression du seuil s de la variable $v \in V^A$. Soit $(\eta^A \rightarrow \eta'^A) \in T^A$ telle que $\eta^A(v) = K_{v, \rho(\eta^A, v)}$. Alors $fold_s^v(\eta^A) \rightarrow fold_s^v(\eta'^A) \in T^B$.*

Démonstration. Deux cas sont à considérer :

1. Si $\eta^A \not\models \Phi_s^v$ ou $\eta'^A \not\models \Phi_s^v$, d'après le Lemme 4, $\eta^A \rightarrow \eta'^A$ est préservée. Autrement dit $fold_s^v(\eta^A) \rightarrow fold_s^v(\eta'^A) \in T^B$.
2. Si $\eta^A \models \Phi_s^v$ et $\eta'^A \models \Phi_s^v$, par hypothèse, la transition $\eta^A \rightarrow \eta'^A$ ne traverse pas le seuil s car en η^A , v n'évolue plus.

— Supposons $\eta^A(v) = \eta'^A(v) = s$.

D'après la définition 11, $fold_s^v(\eta^A)(v) = fold_s^v(\eta'^A)(v) = s - 1$. Pour tout $u \neq v$, $fold_s^v(\eta^A)(u) = \eta^A(u)$ et $fold_s^v(\eta'^A)(u) = \eta'^A(u)$. Par conséquent, $fold_s^v(\eta^A)$ et $fold_s^v(\eta'^A)$ sont adjacents. De plus, nous avons $K_{v, \rho^A(\eta^A, v)}^A = s$ (v n'évolue plus) et $fold_s^v(K_{v, \rho^A(\eta^A, v)}^A) = s - 1$. Il existe un unique $u \neq v$ tel que $\eta^A(u) \neq \eta'^A(u)$. Montrons maintenant que les ressources de $u \in V^A$ ne sont pas affectées par le repliement en η^A . En effet, soit m une ressource de u à l'état η^A . Le repliement de φ_m^A est $fold_s^v((\neg \Phi_s^v \wedge \varphi_m^A) \vee (\Phi_s^v \wedge \varphi_m^A[v \geq s \leftarrow \Phi^{v \geq s}]])$ qui est équivalent à $fold_s^v(\varphi_m^A[v \geq s \leftarrow \Phi^{v \geq s}])$. Ici, $v \geq s$ est vrai et $\Phi^{v \geq s}$ est vrai, car v n'évolue plus. Les autres atomes ne sont pas affectés. m est donc aussi une ressource de u à l'état $fold_s^v(\eta^A)$.

De plus, $fold_s^v$ ne modifie pas les valeurs de $K_{u, \dots}$ pour $u \neq v$.

Puisqu'il y a préservation des ressources et que le repliement ne change pas les positions relatives de η^A , η'^A et $K_{u, \dots}$, nous avons $fold_s^v(\eta^A) \rightarrow fold_s^v(\eta'^A)$.

— Supposons $\eta^A(v) = \eta'^A(v) = s - 1$.

D'après la Définition 11, $fold_s^v(\eta^A)(v) = fold_s^v(\eta'^A)(v) = s - 1$. Pour tout $u \neq v$, $fold_s^v(\eta^A)(u) = \eta^A(u)$ et $fold_s^v(\eta'^A)(u) = \eta'^A(u)$. Par conséquent, $fold_s^v(\eta^A)$ et $fold_s^v(\eta'^A)$ sont adjacents. De plus, nous avons $K_{v, \rho^A(\eta^A, v)}^A = s - 1$ (v n'évolue plus) et $fold_s^v(K_{v, \rho^A(\eta^A, v)}^A) = s - 1$.

Il existe un unique $u \neq v$ tel que $\eta^A(u) \neq \eta'^A(u)$.

Montrons maintenant que les ressources de $u \in V^A$ ne sont pas affectées par le repliement en η^A . En effet, soit m une ressource de u à l'état η^A . Le repliement de φ_m^A est $fold_s^v((\neg \Phi_s^v \wedge \varphi_m^A) \vee (\Phi_s^v \wedge \varphi_m^A[v \geq s \leftarrow \Phi^{v \geq s}]])$ qui est équivalent à $fold_s^v(\varphi_m^A[v \geq s \leftarrow \Phi^{v \geq s}])$. Ici, $v \geq s$ est faux et $\Phi^{v \geq s}$ est faux, car v n'évolue plus. Les autres atomes ne sont pas affectés. m est donc aussi une ressource de u à l'état $fold_s^v(\eta^A)$.

De plus, $fold_s^v$ ne modifie pas les valeurs de $K_{u, \dots}$ pour $u \neq v$.

Puisqu'il y a préservation des ressources et que le repliement ne change pas les positions relatives de η^A , η'^A et $K_{u,\dots}$, nous avons $fold_s^v(\eta^A) \rightarrow fold_s^v(\eta'^A)$.

□

Des lemmes 4 et 5, on peut déduire que la majorité des transitions qui ne traversent pas le seuil sont conservées. Soit l'un des deux états de la transition est loin du seuil et le lemme 4 s'applique ; soit les deux états satisfont Φ_s^v , et si v ne peut plus évoluer, le lemme 5 s'applique. Le lemme 6 traite du dernier cas : pour toutes les transitions $\eta^A \rightarrow \eta'^A$ qui traversent le seuil, les successeurs de $fold_s^v(\eta^A)$ sont les images par $fold_s^v$ des successeurs de η'^A .

Lemme 6. [Nouvelles transitions] Soient $A = (V^A, M^A, E^A, \mathcal{K}^A)$ un RRG, $G^A = (S^A, T^A)$ son graphe d'états, v une variable et s le seuil de v à supprimer. Soient η^A et η'^A deux états de G^A tels que $(\eta^A \rightarrow \eta'^A) \in T^A$ et η^A et η'^A sont de part et d'autre du seuil s . Alors les successeurs de $fold_s^v(\eta^A)$ sont les images par $fold_s^v$ des successeurs de η'^A .

Démonstration. Remarquons que $fold_s^v(\eta^A) = fold_s^v(\eta'^A)$, voir Définition 11. Prouvons maintenant que l'image par $fold_s^v$ de tous les successeurs de η'^A sont les successeurs de $fold_s^v(\eta^A)$.

- Supposons que $\eta^A(v) = s$ et $\eta'^A(v) = s - 1$.
Nous avons $fold_s^v(K_{v,\rho^A(\eta^A,v)}^A) \leq s - 1$, car v diminue, donc $K_{v,\rho^A(\eta^A,v)} \leq s - 1$ et la fonction $fold_s^v$ est l'identité sur $[0 \dots s - 1]$.
 - Les successeurs de η'^A qui ne satisfont pas Φ_s^v sont des successeurs de $fold_s^v(\eta'^A)$, car le lemme 4 est applicable.
 - Les successeurs de η'^A qui satisfont Φ_s^v sont des successeurs de $fold_s^v(\eta'^A)$ car le lemme 5 est applicable. Notons que η^A ne peut pas être successeur de η'^A , car sinon il y aurait une autorégulation négative de v au seuil s .
- Supposons maintenant que $\eta^A(v) = s - 1$ et $\eta'^A(v) = s$.
Nous avons $fold_s^v(K_{v,\rho^A(\eta^A,v)}^A) \geq s - 1$, car v augmente donc $K_{v,\rho^A(\eta^A,v)} \geq s$ et la fonction $fold_s^v$ retranche 1 aux arguments supérieurs ou égaux à s .
 - Les successeurs de η'^A qui ne satisfont pas Φ_s^v sont des successeurs de $fold_s^v(\eta'^A)$, car le lemme 4 est applicable.
 - Les successeurs de η'^A qui satisfont Φ_s^v sont des successeurs de $fold_s^v(\eta'^A)$, car le lemme 5 est applicable. Notons que η^A ne peut pas être successeur de η'^A , car sinon il y aurait une autorégulation de v au seuil s .

Prouvons enfin qu'il n'y a pas d'autres successeurs de $fold_s^v(\eta^A)$. Supposons par l'absurde qu'il existe un successeur dans la direction d'une variable x telle que η^A dans le réseau initial n'a pas de successeur dans cette direction. Dans ce cas, nous avons $fold_s^v(K_{x,\rho^A(\eta^A,x)}^A) \neq fold_s^v(\eta^A)(x)$. Or, pour tout $u \neq v$, $fold_s^v(K_{u,\rho^A(\eta^A,u)}^A) = K_{u,\rho^A(\eta^A,u)}^A$

et $\eta'^A(u) = \text{fold}_s^v(\eta^A)(u)$, donc $K_{x,\rho^A(\eta^A,x)}^A \neq \eta'^A(x)$. Ainsi, dans G^A , il existe une transition sortante de η'^A dans la direction de x , d'où la contradiction. \square

Dans le lemme 6, nous prouvons que, pour une transition qui traverse le seuil, le repliement des successeurs de l'état d'arrivée sont les successeurs du repliement de l'état de départ. Cela exprime bien l'idée de départ de la réduction : elle est basée sur le principe que tout se passe comme si le passage de ce seuil était très rapide et donc que le système devait traverser le seuil avant d'aller vers les autres successeurs.

Le lemme 7 exprime que toutes les transitions dont au moins une des extrémités est non-adjacente au seuil s , proviennent de transitions du réseau initial.

Lemme 7. *Soient $G^A = (S^A, T^A)$ et $G^B = (S^B, T^B)$ les graphes d'états du réseau initial A et du réseau réduit B obtenus par suppression du seuil s de la variable $v \in V^A$. Considérons $(\eta^B \rightarrow \eta'^B) \in T^B$*

1. *Si $\eta^B(v) \neq s - 1$, alors il existe un unique couple (η^A, η'^A) tel que $\eta^B = \text{fold}_s^v(\eta^A)$, $\eta'^B = \text{fold}_s^v(\eta'^A)$ et $(\eta^A \rightarrow \eta'^A) \in T^A$.*
2. *Si $\eta^B(v) = s - 1$, alors il existe un unique triplet $(\eta^A, \eta'^A, \eta''^A)$ tel que $\eta^B = \text{fold}_s^v(\eta^A) = \text{fold}_s^v(\eta'^A)$, $\eta'^B = \text{fold}_s^v(\eta'^A) = \text{fold}_s^v(\eta''^A)$ et on a $\eta^A \rightarrow \eta'^A \rightarrow \eta''^A$.*

Démonstration. Considérons deux états $\eta^B, \eta'^B \in S^B$ tel que $(\eta^B \rightarrow \eta'^B) \in T^B$.

1. Prouvons d'abord la première partie du lemme. D'après le lemme 3, il existe deux états $\eta^A, \eta'^A \in S^A$ tel que $\eta^B = \text{fold}_s^v(\eta^A)$ et $\eta'^B = \text{fold}_s^v(\eta'^A)$. Par définition, il existe un unique u tel que $\eta^B(u) \neq \eta'^B(u)$ et $\eta'^B(u) = \eta^B(u) + \text{sgn}(K_{u,\rho^B(\eta^B,u)}^B - \eta^B(u))$ où $\text{sgn}(x) = 1$ si $x > 0$, $\text{sgn}(x) = 0$ si $x = 0$ et $\text{sgn}(x) = -1$ si $x < 0$.

Prouvons que la transition $\eta^A \rightarrow \eta'^A$ existe.

— Pour $u \neq v$, comme fold_s^v ne change que la variable v , $\eta^B(u) = \eta^A(u)$ et $\eta'^B(u) = \eta'^A(u)$.

— Pour la variable v , 2 cas sont à considérer :

- Si $\eta^B(v) \neq s - 1$:
 - Si $\eta^B(v) < s - 1$ et $\eta'^B(v) < s - 1$, nous avons $\eta^A(v) = \eta^B(v)$ et $\eta'^A(v) = \eta'^B(v)$.
 - Si $\eta^B(v) > \eta'^B(v)$, $K_{v,\rho^B(\eta^B,v)}^B < \eta^B(v)$. D'après la Définition 11, $\text{fold}_s^v(K_{v,\rho^A(\eta^A,v)}^A) = K_{v,\rho^A(\eta^A,v)}^A$. Nous avons $K_{v,\rho^A(\eta^A,v)}^A = K_{v,\rho^B(\eta^B,v)}^B$ alors $K_{v,\rho^A(\eta^A,v)}^A < \eta^A(v)$. La transition $\eta^A \rightarrow \eta'^A$ existe.
 - Si $\eta^B(v) < \eta'^B(v)$, $K_{v,\rho^B(\eta^B,v)}^B > \eta^B(v)$. D'après la Définition 11, soit $K_{v,\rho^A(\eta^A,v)}^A < s - 1$ donc $K_{v,\rho^A(\eta^A,v)}^A = K_{v,\rho^B(\eta^B,v)}^B$ et la transition $\eta^A \rightarrow \eta'^A$ existe, soit $K_{v,\rho^A(\eta^A,v)}^A > s - 1$ donc $K_{v,\rho^A(\eta^A,v)}^A = K_{v,\rho^B(\eta^B,v)}^B + 1$. Dans ce cas $K_{v,\rho^A(\eta^A,v)}^A > \eta^A(v)$ et la transition $\eta^A \rightarrow \eta'^A$ existe.
- Si $\eta^B(v) > s - 1$, nous avons $\eta^A(v) - 1 = \eta^B(v)$ et $\eta'^A(v) - 1 = \eta'^B(v)$.
- Si $\eta^B(v) > \eta'^B(v)$, $K_{v,\rho^B(\eta^B,v)}^B < \eta^B(v)$. D'après la Définition 11, soit $K_{v,\rho^A(\eta^A,v)}^A > s - 1$ donc $K_{v,\rho^A(\eta^A,v)}^A - 1 = K_{v,\rho^B(\eta^B,v)}^B$ dont on déduit

que $K_{v,\rho^A(\eta^A,v)}^A - 1 < \eta^A(v) - 1$ et que la transition $\eta^A \rightarrow \eta'^A$ existe, soit $K_{v,\rho^A(\eta^A,v)}^A < s - 1$ donc $K_{v,\rho^A(\eta^A,v)}^A = K_{v,\rho^B(\eta^B,v)}^B$. Comme $\eta^B(v) > \eta'^B(v) > s - 1$, on a $\eta^B(v) > s$ donc $\eta^A(v) - 1 > s$. Par conséquent, comme $\eta^A(v) - 1 > s > s - 1 > K_{v,\rho^A(\eta^A,v)}^A$, la transition $\eta^A \rightarrow \eta'^A$ existe.

— Si $\eta^B(v) < \eta'^B(v)$, $K_{v,\rho^B(\eta^B,v)}^B > \eta^B(v)$. D'après la Définition 11, $fold_s^v(K_{v,\rho^A(\eta^A,v)}^A) = K_{v,\rho^A(\eta^A,v)}^A - 1$. Nous avons $K_{v,\rho^A(\eta^A,v)}^A - 1 = K_{v,\rho^B(\eta^B,v)}^B$ alors $K_{v,\rho^A(\eta^A,v)}^A - 1 > \eta^A(v) - 1$. La transition $\eta^A \rightarrow \eta'^A$ existe.

— Si $\eta'^B(v) = s - 1$:

— Si $\eta^B(v) = s$, $K_{v,\rho^B(\eta^B,v)}^B < s$. Deux choix sont possibles pour η'^A : $\eta'^A(v) = s$ ou $\eta'^A(v) = s - 1$. Choisissons η'^A tel que $\eta'^A(v) = s$. On a $\eta^A(v) = s + 1$. D'après la Définition 11, $K_{v,\rho^A(\eta^A,v)}^A = K_{v,\rho^B(\eta^B,v)}^B + 1$. Par conséquent, puisque $K_{v,\rho^A(\eta^A,v)}^A < s + 1$, la transition $\eta^A \rightarrow \eta'^A$ existe.

— Si $\eta^B(v) = s - 2$, $K_{v,\rho^B(\eta^B,v)}^B > s - 2$. Deux choix sont possibles pour η'^A : $\eta'^A(v) = s$ ou $\eta'^A(v) = s - 1$. Choisissons η'^A tel que $\eta'^A(v) = s - 1$. On a $\eta^A(v) = s - 2$. Si $K_{v,\rho^B(\eta^B,v)}^B = s - 1$, alors d'après la Définition 11, $K_{v,\rho^A(\eta^A,v)}^A = K_{v,\rho^B(\eta^B,v)}^B$. Par conséquent, puisque $K_{v,\rho^A(\eta^A,v)}^A > s - 2$ et $\eta^A(v) = s - 2$, la transition $\eta^A \rightarrow \eta'^A$ existe. Si $K_{v,\rho^B(\eta^B,v)}^B \geq s$, alors d'après la Définition 11, $K_{v,\rho^A(\eta^A,v)}^A = K_{v,\rho^B(\eta^B,v)}^B + 1$. Par conséquent, puisque $K_{v,\rho^A(\eta^A,v)}^A > s - 2$ et $\eta^A(v) = s - 2$, la transition $\eta^A \rightarrow \eta'^A$ existe.

L'unicité de $(\eta^A, \eta'^A) \in (S^A)^2$ est assurée, car l'un des deux états η^B ou η'^B n'est pas adjacent au seuil s . En effet, les images par $fold_s^v$ des deux états adjacents (la transition $\eta^A \rightarrow \eta'^A$ implique que les deux états soient adjacents), pour lesquels au moins l'un d'eux n'est pas adjacent au seuil s , sont différentes. Ainsi, il ne peut pas exister deux couples différents d'états se reliant sur le couple d'états (η^B, η'^B) .

2. Prouvons maintenant la seconde partie du lemme. Comme $\eta^B = s - 1$, il existe 2 états différents η^A et η'^A tels que $fold_s^v(\eta^A) = fold_s^v(\eta'^A) = \eta^B$. Par définition, il existe un unique u tel que $\eta^B(u) \neq \eta'^B(u)$ et $\eta'^B(u) = \eta^B(u) + \text{sgn}(K_{u,\rho^B(\eta^B,u)}^B - \eta^B(u))$ où $\text{sgn}(x) = 1$ si $x > 0$, $\text{sgn}(x) = 0$ si $x = 0$ et $\text{sgn}(x) = -1$ si $x < 0$. Deux cas sont à considérer.

— Si $u = v$, il existe un unique η''^A tel que $fold_s^v(\eta''^A) = \eta'^B$.

— Si $u \neq v$, posons $\eta''^A = \eta'^A[u \leftarrow \eta'^B(u)]$. Cet état est défini de manière unique une fois que η'^A est fixé.

Montrons que les transitions $\eta^A \rightarrow \eta'^A$ et $\eta'^A \rightarrow \eta''^A$ existent.

— Si $u = v$:

— Si $\eta'^B(v) = s$, $K_{v,\rho^B(\eta^B,v)}^B > s - 1$. On choisit η^A , η'^A et η''^A tels que $\eta^A(v) = s - 1$, $\eta'^A(v) = s$ et $\eta''^A(v) = s + 1$. D'après la Définition 11, $K_{v,\rho^A(\eta^A,v)}^A = K_{v,\rho^B(\eta^B,v)}^B + 1$. Par conséquent, puisque $K_{v,\rho^A(\eta^A,v)}^A > s$, la transition $\eta^A \rightarrow \eta''^A$ existe. Comme les ressources en v sont les mêmes en η^A et en η'^A , on a

- aussi la transition $\eta^A \rightarrow \eta'^A$.
- Si $\eta'^B(v) = s - 2$, $K_{v,\rho^B(\eta^B,v)}^B < s - 1$. On choisit η^A , η'^A et η''^A tels que $\eta^A(v) = s$, $\eta'^A(v) = s - 1$ et $\eta''^A(v) = s - 2$. D'après la Définition 11, $K_{v,\rho^A(\eta'^A,v)}^A = K_{v,\rho^B(\eta^B,v)}^B$. Par conséquent, puisque $K_{v,\rho^A(\eta'^A,v)}^A < s - 1$, la transition $\eta'^A \rightarrow \eta''^A$ existe. Comme les ressources en v sont les mêmes en η^A et en η'^A on a aussi la transition $\eta^A \rightarrow \eta'^A$.
 - Si $u \neq v$: soient η et η' les deux états de S^A tels que $fold_s^v(\eta) = fold_s^v(\eta') = \eta^B$. On a l'une des transitions suivantes $\eta \rightarrow \eta'$ ou $\eta' \rightarrow \eta$, car sinon, il y aurait une autorégulation de v au seuil s . Définissons η^A comme étant la source de cette transition et η'^A comme la destination. On a donc $\eta^A \rightarrow \eta'^A \in T^A$. Sans perte de généralité, on considère que $K_{u,\rho(\eta^B,u)}^B > \eta^B(u)$. Comme $fold_s^v(K_{u,\dots}^B) = K_{u,\dots}^B = K_{u,\rho(\eta^B,u)}^B$, et $\eta^A(u) = \eta^B(u)$ et $\eta'^A(u) = \eta'^B(u)$, on a la transition $\eta^A \rightarrow \eta'^A$.

□

Théorème 1. [Préservation des états stables] Soient $A = (V^A, M^A, E^A, \mathcal{K}^A)$ un RRG, $B = (V^B, M^B, E^B, \mathcal{K}^B)$ le réseau réduit obtenu par suppression d'un seuil s de la variable $v \in V^A$, η^A un état de A et $\eta^B = fold_s^v(\eta^A)$. Soit η'^A le symétrique de η^A par rapport au seuil s .

1. Si η^A est un état stable de A alors η^B est stable dans B .
2. Si η^B est stable et $\eta^B(v) \neq s - 1$, alors η^A est stable.
3. Si η^B est stable et $\eta^B(v) = s - 1$, alors soit η^A soit η'^A est stable.

Démonstration. 1. Prouvons d'abord l'item 1 du théorème. Deux cas sont à considérer :

- Si $\eta^A \notin \Phi_s^v$ (η^A est loin du seuil à supprimer), le lemme 2 assure que les ressources sont préservées par la réduction. Supposons que η^A est un état stable, autrement dit $\forall u \in V^A, \eta^A(u) = K_{u,\rho^A(\eta^A,u)}^A$.
- Si $\eta^A(v) < s - 1$, pour tout $u \in V^A$, nous avons $fold_s^v(K_{u,\rho^A(\eta^A,u)}^A) = K_{u,\rho^A(\eta^A,u)}^A$ et $fold_s^v(\eta^A(u)) = \eta^A(u)$. Donc, $fold_s^v(K_{u,\rho^A(\eta^A,u)}^A) = fold_s^v(\eta^A(u))$ et par conséquent, η^B est stable.
- Si $\eta^A(v) > s$, pour tout $u \neq v$, $fold_s^v$ ne change ni la valeur de $\eta^A(u)$, ni les paramètres $K_{u,\dots}$ de u , donc $fold_s^v(K_{u,\rho^A(\eta^A,u)}^A) = K_{u,\rho^A(\eta^A,u)}^A$. Pour la variable variable v , nous avons $fold_s^v(\eta^A(v)) = \eta^A(v) - 1$ et $fold_s^v(K_{v,\rho^A(\eta^A,v)}^A) = K_{v,\rho^A(\eta^A,v)}^A - 1$. Par conséquent, η^B est un état stable.
- Si $\eta^A \in \Phi_s^v$ (η^A est adjacent au seuil à supprimer), quelque soit la valeur de $\eta^A(v) \in \{s - 1, s\}$, l'une des deux transitions $\eta^A \rightarrow \eta'^A$ ou $\eta'^A \rightarrow \eta^A$ existe, sinon il y aurait une autorégulation sur le seuil s de v . Si η^A est stable, il ne peut pas y avoir de transition sortante. Par conséquent, la transition $\eta'^A \rightarrow \eta^A$ existe. Le lemme 6 est alors applicable : les successeurs de $fold_s^v(\eta'^A) = fold_s^v(\eta^A) = \eta^B$ sont les images par $fold_s^v$ des successeurs de η^A . η^A n'ayant pas de successeurs, η^B n'en a pas non plus. η^B est donc stable.

2. Prouvons maintenant l'item 2 du théorème. Considérons η^B un état stable du réseau réduit tel que $\eta^B \neq s - 1$. Comme η^B est stable, $\forall u \in V^B$, $K_{u,\rho(\eta^B,u)}^B = \eta^B(u)$.
- Pour $u \neq v$, on a $fold_s^v(K_{u,\rho(\eta^A,u)}^A) = fold_s^v(\eta^A(u))$. Comme la fonction $fold_s^v$ ne modifie pas les paramètres $K_{u,\dots}$ et les composantes u des états, on en déduit que $K_{u,\rho(\eta^A,u)}^A = \eta^A(u)$ et donc que la variable u est stable en η^A .
 - Pour v , deux cas sont à considérer.
 - Si $\eta^B(v) < s - 1$, on a $fold_s^v(K_{v,\rho(\eta^A,v)}^A) = fold_s^v(\eta^A(v))$. Comme la fonction $fold_s^v$ est l'identité sur $[0 \dots s - 1]$, on en déduit que $K_{v,\rho(\eta^A,v)}^A = \eta^A(v)$ et donc que la variable v est stable en η^A .
 - Si $\eta^B(v) > s - 1$, on a $fold_s^v(K_{v,\rho(\eta^A,v)}^A) + 1 = fold_s^v(\eta^A(v)) + 1$. On en déduit que $K_{v,\rho(\eta^A,v)}^A = \eta^A(v)$ et donc que la variable v est stable en η^A .
3. Prouvons enfin l'item 3 du théorème. Considérons η^B un état stable du réseau réduit tel que $\eta^B = s - 1$. Comme η^B est stable, $\forall u \in V^B$, $K_{u,\rho(\eta^B,u)}^B = \eta^B(u)$. Une seule des deux transitions $\eta^A \rightarrow \eta'^A$ ou $\eta'^A \rightarrow \eta^A$ existe, sinon il y aurait une autorégulation sur le seuil s de v . Supposons, sans perte de généralité, que ce soit la transition $\eta^A \rightarrow \eta'^A$. D'après le lemme 5, les successeurs de $\eta^B = fold_s^v(\eta^A)$ sont les images par $fold_s^v$ des successeurs de η'^A . Supposons par l'absurde que η'^A ait un successeur, η^B aurait également un successeur : contradiction.

□

Nous terminons cette section en donnant le théorème de préservation faible des attracteurs cycliques.

Théorème 2. [Préservation faible des attracteurs cycliques] Soient $G^A = (S^A, T^A)$ et $G^B = (S^B, T^B)$ les graphes d'états du réseau initial A et du réseau réduit B obtenus par suppression du seuil s de la variable v , et $\eta, \eta' \in S^A$ deux états tels que $\eta \rightarrow \eta' \in T^A$.

- Si $\eta \rightarrow \eta'$ est une transition d'un cycle de G^A duquel il n'est pas possible de sortir, alors $fold_s^v(\eta) \rightarrow fold_s^v(\eta')$ est une transition d'un cycle de G^B .
- Si $fold_s^v(\eta) \rightarrow fold_s^v(\eta')$ est une transition d'un cycle de G^B duquel il n'est pas possible de sortir, alors $\eta \rightarrow \eta'$ est une transition d'un cycle de G^A .

Démonstration. Considérons un cycle $\eta_1 \rightarrow \eta_2 \rightarrow \eta_3 \rightarrow \dots \rightarrow \eta_{2n-1} \rightarrow \eta_{2n} \rightarrow \eta_1$ dans le graphe d'états G^A duquel il n'est pas possible de sortir.

1. Si η_i ou η_{i+1} ne satisfait pas Φ_s^v , le lemme 4 est applicable et implique la préservation de la transition.
2. Si η_i et η_{i+1} satisfont Φ_s^v et s'ils sont du même côté du seuil s , le lemme 5 s'applique, parce que la seule variable qui peut évoluer est celle qui fait passer de η_i à η_{i+1} (v ne peut pas bouger). Le lemme 5 implique que $(fold_s^v(\eta_i) \rightarrow fold_s^v(\eta_{i+1})) \in T^B$.
3. Si η_i et η_{i+1} satisfont Φ_s^v et s'ils sont de part et d'autre du seuil :
 - nous avons $fold_s^v(\eta_i) = fold_s^v(\eta_{i+1})$.

— le lemme 6 montre que les successeurs de $fold_s^v(\eta_i)$ sont des image par $fold_s^v$ des successeurs de η_{i+1} . Comme on peut pas sortir du cycle, il n'y a qu'un seul successeur de $\eta_{i+1} : \eta_{i+2}$ (où les indices $i + 1$ et $i + 2$ est pris modulo $2n$ parce qu'il s'agit d'un cycle de longueur $2n$), donc le seul successeur de $fold_s^v(\eta_i)$ est $fold_s^v(\eta_{i+2})$.

On a montré que l'image par $fold_s^v$ de η_i n'a qu'un seul successeur. Puisque l'ensemble des états S^B est fini, chaque état $fold_s^v(\eta_i)$ appartient à un cycle.

Réciproquement, considérons un cycle $\eta_1^B \rightarrow \eta_2^B \rightarrow \eta_3^B \rightarrow \dots \rightarrow \eta_{2n-1}^B \rightarrow \eta_{2n}^B \rightarrow \eta_1^B$ dans le graphe d'états G^B duquel il n'est pas possible de sortir.

1. Supposons que pour tout état du cycle, $\eta_i^B(v) \neq s - 1$. Pour toutes les transitions du cycle, le lemme 7-1 s'applique et il existe donc un cycle dans A qui explique la présence du cycle dans B .
2. Supposons que si $\eta_i^B(v) = s - 1$, alors $\eta_{i-1}^B = s - 1$ et $\eta_{i+1}^B = s + 1$ (les indices sont calculés modulo $2n$). Pour chacune des transitions, on peut appliquer soit le lemme 7-1, soit le lemme 7-2. On construit donc pas à pas, un cycle dans le réseau A qui se replie sur le cycle considéré dans B .
3. Supposons que le chemin $\eta_i^B \rightarrow \dots \rightarrow \eta_k^B$ est tel que $\eta_i^B(v) \neq s - 1$, $\eta_k^B(v) \neq s - 1$ et $\forall j \in]i, k[$, $\eta_j^B = s - 1$.

La première transition $\eta_i^B \rightarrow \eta_{i+1}^B$ permet de construire, *via* le lemme 7-1, une unique transition η_i^A vers η_{i+1}^A .

Pour chaque $\forall j \in]i, k[$, on peut associer à η_j^B , 2 états η_j et η'_j qui se replient tous les deux en η_j^B . On a une et une seule des transitions suivantes : $\eta_j \rightarrow \eta'_j$ ou $\eta_j \rightarrow \eta'_j$. Posons η_j^A la source de cette transition et η'_j^A la destination. On a donc $\eta_j^A \rightarrow \eta'_j^A$. Par le lemme 6, on a une transition de η'_j^A vers $\eta''_j^A = \eta_j^A[u \leftarrow \eta'^B(u)]$, où u est la seule variable qui change entre η_i^B et η_{i+1}^B .

La dernière transition $\eta_{k-1}^B \rightarrow \eta_k^B$ permet de construire, *via* le lemme 7-2, les deux transitions η_{k-1}^A vers η_{k-1}^A et η_{k-1}^A vers η_{k-1}^A .

Il reste à montrer qu'il existe un chemin de η_i^A vers η_k^A . Par induction :

- $\eta_i^A \rightarrow \eta_{i+1}^A$ permet d'entrer dans l'ensemble $\{\eta_{i+1}^A, \eta'_{i+1}^A\}$.
- Supposons qu'il existe un chemin qui mène de η_i^A à $\{\eta_j^A, \eta'_j^A\}$. Alors, on peut construire un chemin de η_i^A à l'ensemble $\{\eta_{j+1}^A, \eta'_{j+1}^A\}$. En effet, si le premier chemin arrive en η_j^A , alors on le continue par $\eta_j^A \rightarrow \eta'_j^A$ puis par $\eta'_j^A \rightarrow \eta''_j^A$ qui appartient à $\{\eta_{j+1}^A, \eta'_{j+1}^A\}$.
- La dernière transition permet de construire un chemin de η_i^A à η_k^A .

Pour construire le cycle dans G^A il suffit de répéter, la même démarche jusqu'à retomber sur un état déjà parcouru (le nombre d'états est fini).

4. Supposons que $\forall i \in [1, 2n]$, $\eta_i^B(v) = s - 1$ (en particulier, c'est le cas que l'on applique lorsqu'on a supprimé le dernier seuil d'une variable). Le cycle s'écrit $\eta_1^B \rightarrow \dots \rightarrow \eta_{2n}^B \rightarrow \eta_1^B$. Le principe d'induction du cas précédent mène à l'existence d'un cycle commençant en η_1^A .

□

En accord avec les résultats démontrés dans ce chapitre, dans l'exemple du chapitre 2 (section 2.3, page 39), nous avons vu que dans les deux réductions, les comportements asymptotiques sont conservés. L'état stable $(0, 1)$ du réseau initial devient, dans le réseau réduit, soit l'état stable (0) lorsque l'on supprime u , soit l'état stable $(0, 1)$ lorsqu'on supprime le seuil 1 de v . Le cycle de taille 2 entre les états $(1, 0)$ et $(2, 0)$ du réseau initial est également conservé : après la suppression de la variable u , il est présent entre les états (1) et (2) , et après la suppression du seuil 1 de v , il est présent entre les états $(0, 0)$ et $(1, 0)$.

En conclusion, nous avons montré en toute généralité que chaque suppression de seuil d'une variable préserve les attracteurs cycliques et les états stables.

Chapitre 4

Application de la réduction au phage λ

Ce chapitre est dédié à l'application des réductions définies dans les chapitres précédents, à la modélisation du switch génétique du phage λ . Il s'agit d'un mécanisme largement étudié qui reste intéressant, car son réseau est suffisamment simple pour l'appréhender facilement et son comportement (switch génétique) est suffisamment complexe pour que sa modélisation n'en soit pas triviale. Il constitue donc un très bon exemple sur lequel nous pouvons appliquer la méthode de réduction.

Dans ce chapitre, nous montrons la cohérence entre deux modèles du phage λ : l'un à 4 variables et l'autre à 2 variables, tous deux construits indépendamment à partir de données biologiques. Nous réduisons successivement le réseau à 4 variables et finissons par obtenir celui à 2 variables. Enfin, nous montrons que nous pouvons aller plus loin dans la réduction du modèle en atteignant un réseau booléen irréductible.

4.1 Le mécanisme du switch chez le phage λ

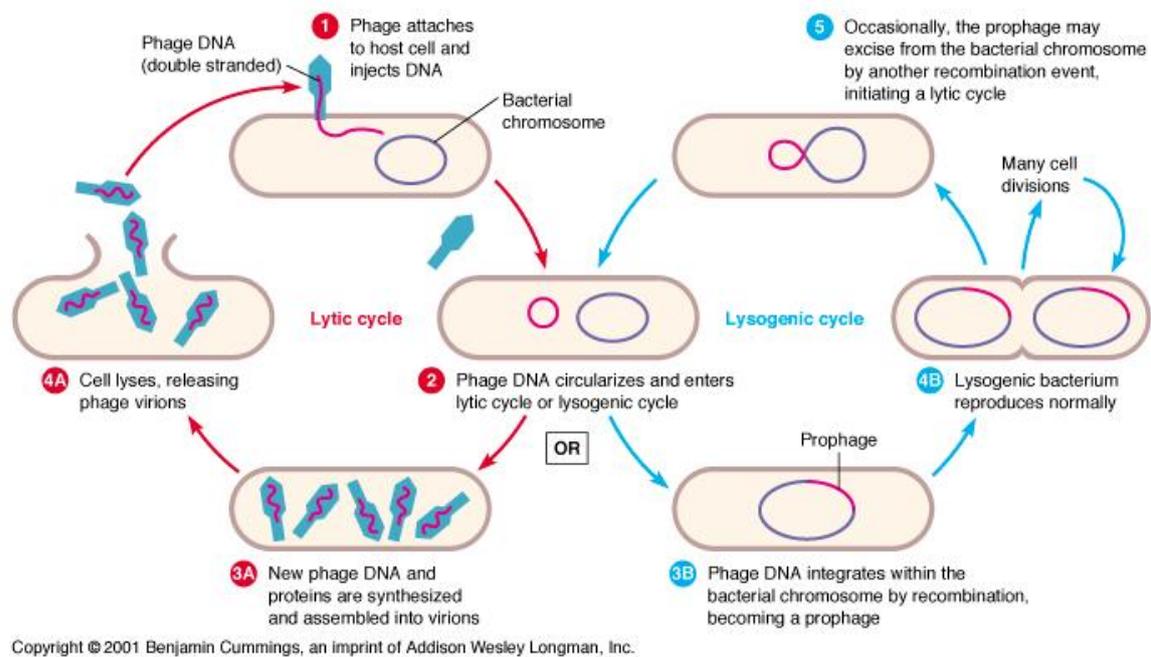
L'*Enterobacteria* λ phage est un virus infectant la bactérie *Escherichia Coli* qui est présente dans la partie basse de l'intestin de nombreux organismes. Le virus contient un double brin d'ADN qui lui est spécifique, et qui contient les gènes permettant sa reproduction. Comme tous les virus, il se sert de la machinerie de la cellule infectée pour se répliquer en très grand nombre afin d'infecter d'autres cellules.

La particularité de ce virus est la « prise de décision » qu'il effectue lors de l'infection, en fonction de l'état de la cellule. Ce mécanisme est dû à un « switch génétique ¹ » qui est depuis longtemps très bien décrit [41, 29, 28, 57]. De manière générale, un switch génétique est un système qui, selon un stimulus extérieur, bascule d'un état vers un autre. Dans le cas du phage λ , ces états sont la *lyse* et la *lysogénie* (voir Figure 4.1).

- La lyse est l'état de la cellule qui permet au virus de se répliquer en grande quantité grâce à la machinerie cellulaire. Cette phase se finit invariablement par la destruction de la cellule et la libération des virus. Le brin d'ADN viral est injecté par le virus dans la cellule, il est alors répliqué et transcrit dans la cellule afin de synthétiser les éléments du virus.
- La lysogénie est la phase dormante du virus. Dans certaines conditions, l'ADN viral est intégré à l'ADN bactérien. Il va donc être répliqué lors de la division des bactéries et, selon le temps qu'il va passer en dormance, il va se répandre dans la population bactérienne. Quand il est dans cet état, l'ADN viral peut s'exciser de l'ADN bactérien. Une fois l'ADN libéré, la cellule passe en phase lytique.

Les stimuli qui pilotent cette prise de décision ne sont pas précisément connus. Cependant, nous savons que le basculement vers la lysogénie est induit par de mauvaises conditions de croissance de la bactérie (manque de nutriments, températures basses, infec-

1. Le terme de « switch génétique » est souvent utilisé dans la littérature, mais il s'agit d'un switch épigénétique, car c'est un élément environnemental le fait changer d'état, et non pas une mutation.

FIGURE 4.1 – Schéma représentant le switch génétique du phage λ .

tions multiples) [56]. Dans de mauvaises conditions, le virus « attend » un environnement plus riche pour optimiser la synthèse de nouveaux virus.

Le switch génétique du phage λ est sous le contrôle d'un réseau génétique largement étudié [58, 29, 28]. Il contient 4 principaux gènes : *CI*, *Cro*, *CII* et *N*. Les phases lysogénique et lytique sont caractérisées par les niveaux des protéines des gènes *CI* et *Cro* :

- La phase lysogénique est associée à un niveau haut de protéines *CI* et un niveau bas de protéines *CRO*.
- La phase lytique est associée à un niveau bas de protéines *CI* et un niveau haut de protéines *CRO*.

De plus, les régulations entre les 4 gènes précités se résument par :

- *CI* inhibe les gènes *Cro*, *CII* et *N*. Cependant, les produits du gène *CI* l'activent au même niveau de concentration que celui auquel ils inhibent *Cro*.
- *Cro* inhibe le gène *CI*. Cependant, à une haute concentration du produit de *Cro*, *Cro* s'inhibe lui-même.
- *CII* active le gène *CI*.
- *N* active le gène *CII*.

Le nombre de variables ne permet pas de connaître la dynamique intuitivement (nombre de jeux de paramètres possibles trop important) et la modélisation est un outil extrêmement utile à la biologie pour mieux comprendre le comportement de tels switches au niveau moléculaire. Les expérimentations montrent qu'il y a une prise de décision qui est faite et qui fait basculer le réseau vers l'une des deux phases.

Du point de vue génétique, la clé du switch réside dans la protéine *CII* : dans un

environnement riche en nutriments, la machinerie cellulaire va dégrader cette protéine. CI ne sera pas activé et le gène Cro ne sera donc pas inhibé. Le système va donc basculer en lyse. À l'inverse, quand le milieu est pauvre en nutriments, la cellule est au ralenti et elle ne peut pas dégrader CII . C'est donc CI qui va prendre le dessus sur Cro et la cellule passe en lysogénique.

Après une phase lysogénique, l'excision de l'ADN viral est provoquée par la présence de 2 protéines (xis et int). Une fois l'ADN viral libéré, la phase lytique est lancée.

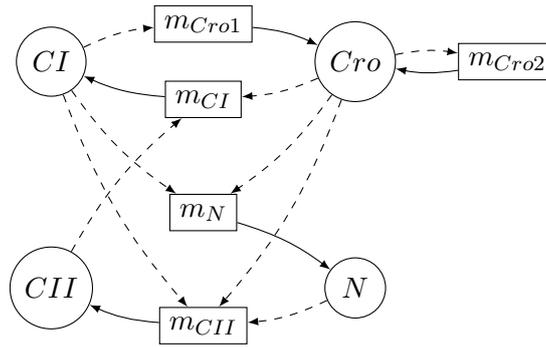
D'assez nombreux modèles du réseau génétique du phage λ existent :

- Gary Arcker *et al.* [1] ont proposé un modèle différentiel basé sur les gènes CI et Cro pour montrer l'importance de l'interaction entre ces deux gènes dans la dynamique du switch.
- Tianhai Tian *et al.* [72] ont proposé des simulations stochastiques afin d'étudier la bistabilité du switch.
- Denis Mestivier *et al.* [52] ont présenté plusieurs approches différentes pour modéliser le switch du phage λ (quantitatives, stochastiques et qualitatives) et les ont comparées. Nous utiliserons par la suite les valeurs des paramètres donnés dans le chapitre consacré à la modélisation qualitative.
- Adam Arkin *et al.* [6] ont développé une approche stochastique pour modéliser le switch du phage λ .
- Karsten R. Heidtke *et al.* [39] ont développé un modèle qualitatif, basé sur l'environnement de simulation QSim [46].

Le choix du nombre de variables d'un modèle impacte la capacité à représenter des comportements biologiques détaillés. Par exemple, un modèle avec peu de variables peut être une abstraction d'un autre modèle en contenant davantage. Il y a donc un réel intérêt à établir et automatiser l'identification des relations d'abstraction entre ces différents modèles afin d'établir leur éventuelle compatibilité. Dans la suite du chapitre, nous allons partir d'un modèle avec un nombre important de variables, puis effectuer des réductions successives, ce qui nous permettra de trouver un modèle plus petit ayant les mêmes comportements asymptotiques que celui d'origine.

4.2 Deux modèles classiques du phage

Le modèle initial que nous allons utiliser pour illustrer la réduction contient les 4 gènes principaux du switch du phage λ (CI , Cro , CII et N). Ce modèle a été introduit par Denis Thieffry et René Thomas en 1995 [67] pour représenter et modéliser un ensemble d'informations expérimentales complexes [68]. La figure 4.2 contient le graphe d'interaction (en haut) et le réseau avec ses différents éléments (en bas). Les valeurs des paramètres choisis dans cette figure ne représentent qu'un seul jeu de paramètres parmi 88 compatibles avec les caractéristiques dynamiques du phage λ .

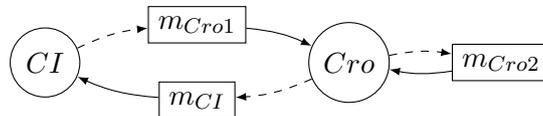


$V :$	$M :$	$E :$
Cro ($b_{Cro} = 3$)	$m_{Cro1} : \neg CI \geq 1$	$(m_{Cro1} \rightarrow Cro)$
CI ($b_{CI} = 2$)	$m_{Cro2} : \neg Cro \geq 1$	$(m_{Cro2} \rightarrow Cro)$
CII ($b_{CII} = 1$)	$m_{CI} : \neg Cro \geq 1 \vee CII \geq 1$	$(m_{CI} \rightarrow CI)$
N ($b_N = 1$)	$m_{CII} : \neg(CI \geq 2) \wedge \neg(Cro \geq 3) \wedge (N \geq 1)$	$(m_{CII} \rightarrow CII)$
	$m_N : \neg(CI \geq 1) \wedge \neg(Cro \geq 2)$	$(m_N \rightarrow N)$

$\mathcal{K} :$			
$K_{Cro} = 0$	$K_{CI} = 0$	$K_{CII} = 0$	$K_N = 0$
$K_{Cro, \{m_{Cro1}\}} = 2$	$K_{CI, \{m_{CI}\}} = 2$	$K_{CII, \{m_{CII}\}} = 1$	$K_{N, \{m_N\}} = 1$
$K_{Cro, \{m_{Cro2}\}} = 0$			
$K_{Cro, \{m_{Cro1}, m_{Cro2}\}} = 3$			

 FIGURE 4.2 – Modèle du réseau génétique à 4 variables responsable du switch chez le phage λ

Le second modèle (le « petit » modèle) est celui utilisé comme exemple dans la section 2.2. Il est présenté dans le même article [67] qui n'établit toutefois pas de liens rigoureux entre les deux modèles. La figure 4.3 contient le graphe d'interaction (en haut) et le réseau avec ses différents éléments (en bas).



$V :$	$M :$	$E :$	$\mathcal{K} :$
Cro ($b_{Cro} = 2$)	$m_{Cro1} : \neg(CI \geq 1)$	$(m_{Cro1} \rightarrow Cro)$	$K_{Cro} = 0$
CI ($b_{CI} = 1$)	$m_{Cro2} : \neg(Cro \geq 2)$	$(m_{Cro2} \rightarrow Cro)$	$K_{CI} = 0$
	$m_{CI} : \neg(Cro \geq 1)$	$(m_{CI} \rightarrow CI)$	$K_{CI, \{m_{CI}\}} = 1$
			$K_{Cro, \{m_{Cro2}\}} = 0$
			$K_{Cro, \{m_{Cro1}, m_{Cro2}\}} = 2$

 FIGURE 4.3 – Réseau de régulation génétique à 2 variables responsable du switch du phage λ .

4.3 Réduction du modèle à 4 variables

La complexité des opérations d'une réduction m'a amené à développer un script permettant de réduire automatiquement un réseau génétique sur un seuil d'une variable donnée. Ce programme est implémenté en Python. L'entrée du programme est un fichier décrivant le réseau, au format SMBioNet [44] qui contient plusieurs parties décrivant les variables, les multiplexes et leurs arcs sortants, et enfin les paramètres.

La première étape consiste à convertir le fichier d'entrée décrivant le réseau, en informations stockées dans la structure interne du programme. En particulier, les formules des multiplexes doivent être transformées en structures arborescentes faciles à manipuler lors des réductions successives du réseau. L'étape suivante demande à l'utilisateur le nom d'une variable et un de ses seuils et calcule le réseau réduit. La modification des formules des multiplexes mène à des formules de grande taille (combinatoire des ressources) qu'il faut essayer de simplifier à la volée. La stratégie utilisée ici est de descendre les négations au niveau des feuilles, de chercher les tautologies et antilogies et de simplifier l'arbre en conséquence.

Ce programme a été utilisé pour obtenir les réductions qui suivent.

Dans toute la sous-section, nous nommerons tous les états par un n -uplet dans lequel la position d'une valeur indique la variable concernée.

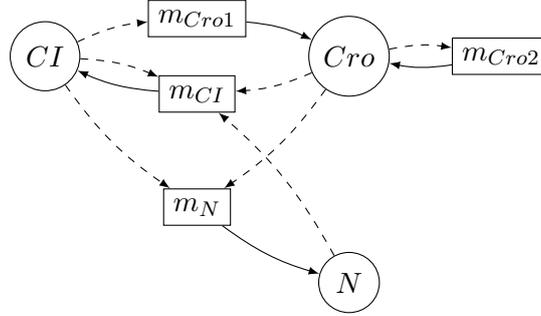
- Pour le modèle à 4 variables, les n -uplets seront donnés dans l'ordre (CI, Cro, CII, N) . Ainsi, l'état $(0, 2, 1, 0)$ signifie que CI et N ont la valeur 0, Cro à la valeur 2 et CII à la valeur 1.
- Pour le modèle à 3 variables (CII ayant été supprimé), les n -uplets seront donnés dans l'ordre (CI, Cro, N) . Ainsi, l'état $(0, 2, 1)$ signifie que CI à la valeur 0, Cro la valeur 2 et N la valeur 1.
- Pour le modèle à 2 variables (CII et N ayant été supprimés), les n -uplets seront donnés dans l'ordre (CI, Cro) . Ainsi, l'état $(2, 1)$ signifie que $CI = 2$ et $Cro = 1$.

Le réseau de départ (Figure 4.2) contient 48 états. La dynamique contient 2 bassins d'attractions : l'état stable $(2, 0, 0, 0)$ et le cycle de taille 2 entre les états $(0, 2, 0, 0)$ et $(0, 3, 0, 0)$. Ce sont donc les deux éléments de la dynamique qui seront conservés au cours des réductions.

Nous allons maintenant montrer comment le réseau à 4 variables peut être transformé *via* une suite de réductions pour donner le réseau à 2 variables. Les réductions successives seront : suppression de la variable booléenne CII , suppression de la variable booléenne N , suppression du seuil 2 de la variable Cro , et enfin suppression du seuil 1 de CI . Une 5^{ème} réduction est encore possible (suppression du seuil 1 de Cro) menant à un réseau à 2 variables ne contenant que 4 états.

4.3.1 Suppression de la variable booléenne CII

Le seul multiplexe qui sera affecté par cette réduction est m_{CI} , car il contient l'atome $CII \geq 1$. En remplaçant cet atome par la formule indiquant si les prédécesseurs de CII lui permettent de passer son seuil et en effectuant le repliement des états des formules et des paramètres ($fold_s^v$), on obtient le réseau réduit de la figure 4.4.



$V :$	$M :$	$E :$
Cro ($b_{Cro} = 3$)	$m_{Cro1} :$ $\neg CI \geq 2$	$(m_{Cro1} \rightarrow Cro)$
CI ($b_{CI} = 2$)	$m_{Cro2} :$ $\neg Cro \geq 3$	$(m_{Cro2} \rightarrow Cro)$
N ($b_N = 1$)	$m_{CI} :$ $\neg(Cro \geq 1) \vee (\neg(Cro \geq 3) \wedge \neg(CI \geq 2) \wedge (N \geq 1))$	$(m_{CI} \rightarrow CI)$
	$m_N :$ $\neg(CI \geq 1) \wedge \neg(Cro \geq 2)$	$(m_N \rightarrow N)$

$\mathcal{K} :$		
$K_{Cro} = 0$	$K_{CI} = 0$	$K_N = 0$
$K_{Cro, \{m_{Cro1}\}} = 2$	$K_{CI, \{m_{CI}\}} = 2$	$K_{N, \{m_N\}} = 1$
$K_{Cro, \{m_{Cro2}\}} = 0$		
$K_{Cro, \{m_{Cro1}, m_{Cro2}\}} = 3$		

FIGURE 4.4 – Réseau génétique obtenu par suppression de la variable CII à partir du réseau de la figure 4.2

4.3.2 Suppression de la variable booléenne N

La seule formule de multiplexe contenant l'atome $(N \geq 1)$ est $\varphi_{m_{CI}}$. On remplace donc cet atome par la formule $\Phi^{v \geq 1}$, puis on replie l'ensemble du réseau. On obtient le réseau de la figure 4.5.

Le réseau ne contient maintenant plus que 2 variables : Cro et CI . La dynamique de ce nouveau réseau (décrit dans la figure 4.5) est représentée par le graphe d'états de la figure 4.6. On y retrouve les deux comportements asymptotiques présentés plus haut : un état stable (l'état $(0, 2)$) et un cycle de longueur 2 (entre les états $(2, 0)$ et $(3, 0)$). Il y a d'autres cycles de longueur 2 : un sur le seuil 1 de CI et un autre sur le seuil 3 de Cro . Cependant, ces deux derniers cycles ne sont pas des comportements asymptotiques, car il y a des transitions qui permettent au système de s'en échapper.

La méthode de réduction de Naldi ne permet pas de réduire davantage le réseau, car

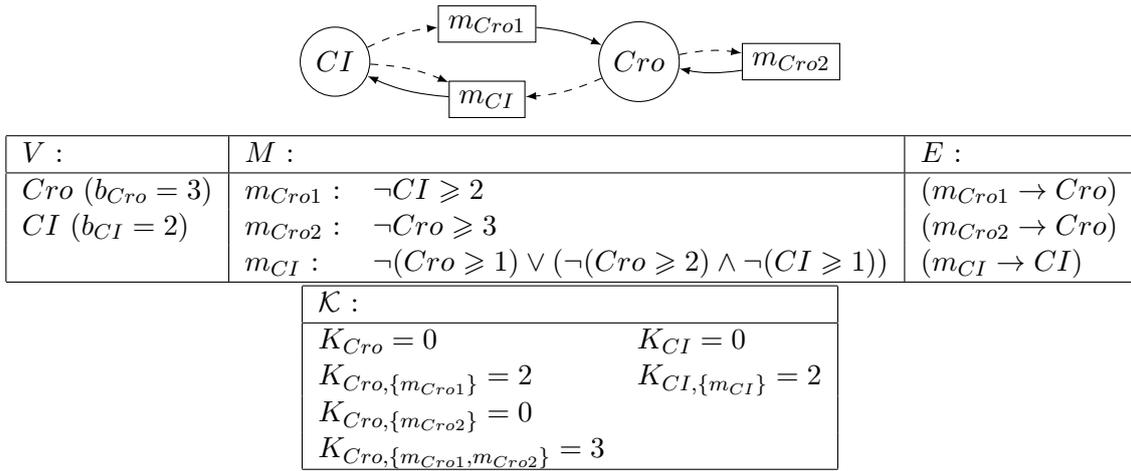


FIGURE 4.5 – Réseau génétique obtenu par suppression de la variable N à partir du réseau de la figure 4.4

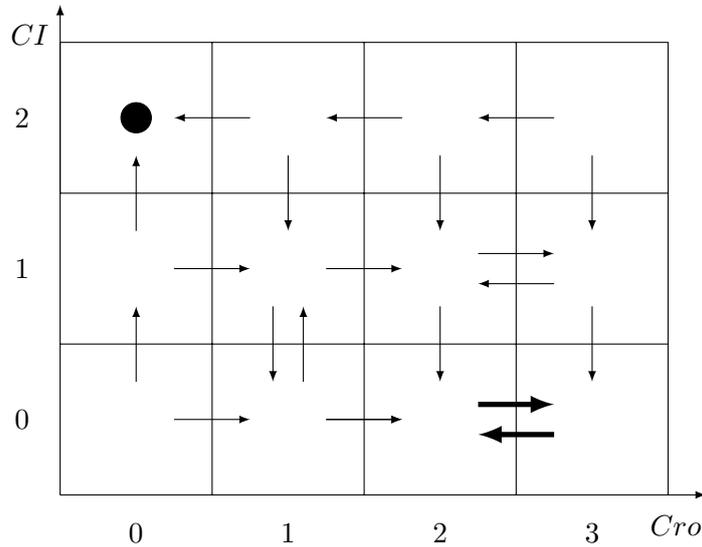


FIGURE 4.6 – Dynamique du réseau de la figure 4.5

il y a des autorégulations négatives sur chacune des variables. Pour la variable Cro , cette autorégulation se voit dans le multiplexe m_{Cro2} dont la formule contient le terme Cro et dont le seul successeur est Cro . Pour la variable CI , elle se voit dans le multiplexe m_{CI} dont la formule contient le terme CI et dont le seul successeur est CI (voir Figure 4.5). Notre méthode permet d'aller plus loin dans la réduction.

4.3.3 Suppression du seuil 2 de Cro

Pour la suppression du seuil 2 de Cro , un seul multiplexe possède une formule qui contient l'atome $Cro \geq 2$. Cet atome est donc remplacé par $\Phi^{v \geq 1}$. Le réseau obtenu est représenté dans la figure 4.7.

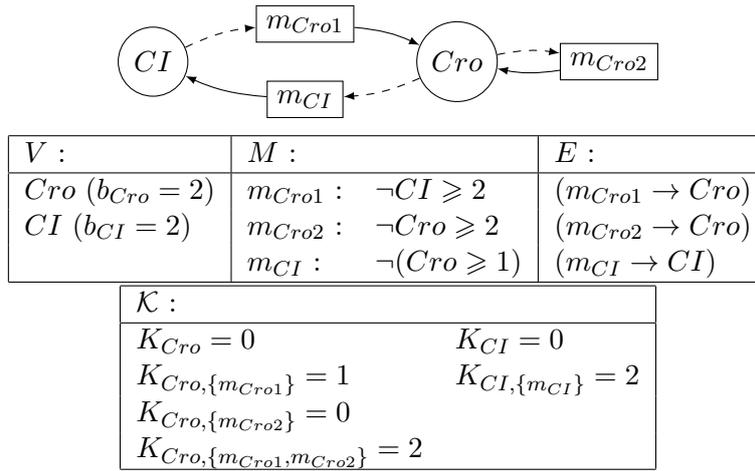


FIGURE 4.7 – Réseau génétique obtenu par suppression du seuil 2 de la variable Cro à partir du réseau de la figure 4.5

La figure 4.8 représente la dynamique de ce réseau. Nous retrouvons bien l'état stable et le cycle de longueur 2 (en épais). Parmi les autres cycles non asymptotiques, celui sur CI a disparu, ce qui n'est pas surprenant, car il n'a pas de raison d'être conservé pendant une réduction.

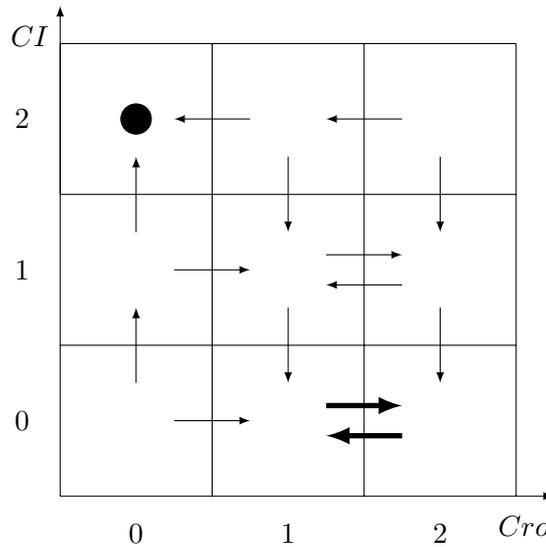


FIGURE 4.8 – Dynamique du réseau de la figure 4.7

4.3.4 Suppression du seuil 1 de CI

Il n'y a aucun atome dont la formule contient l'atome $CI \geq 1$. Il faut donc simplement effectuer un repliement du réseau. Le réseau obtenu est celui de la figure 4.9 et sa dynamique présentée en Figure 4.10.

Le réseau obtenu est le réseau avec 2 variables présenté dans la section 2.2. Avec notre

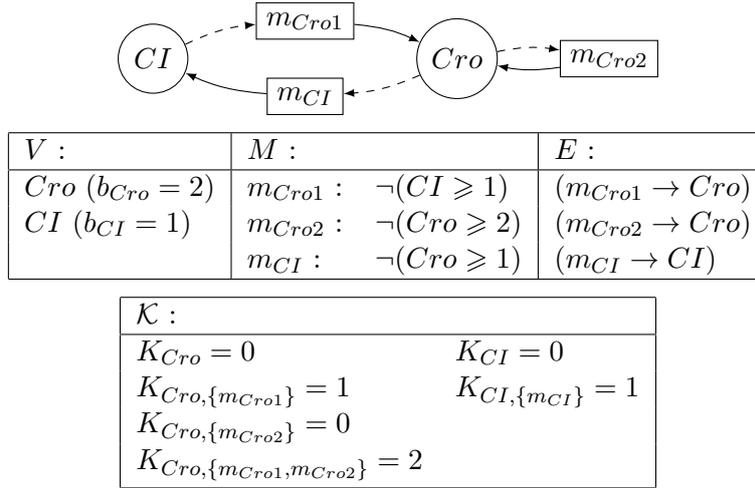


FIGURE 4.9 – Réseau génétique obtenu par suppression du seuil 1 de CI à partir du réseau de la figure 4.7

méthode de réduction de réseau, nous avons donc réduit le modèle contenant 4 variables et obtenu le réseau à 2 variables, les deux ayant été validés d'un point de vue biologique. Il y a donc une relation d'abstraction : le modèle à 2 variables est une abstraction du modèle à 4 variables par rapport aux comportements asymptotiques.

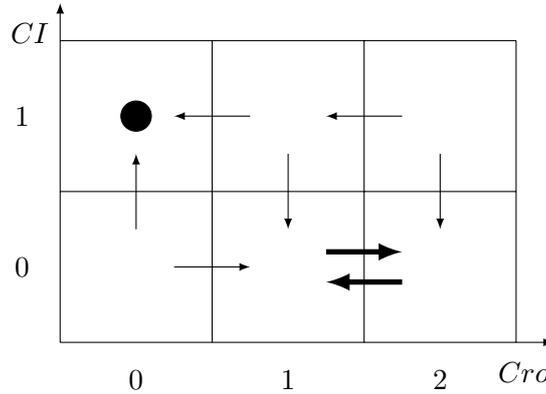


FIGURE 4.10 – Dynamique du réseau de la figure 4.9

4.3.5 Suppression du seuil 1 de Cro

Nous pouvons encore réduire le réseau précédent en supprimant le seuil 1 de Cro . Le seul multiplexe à modifier est m_{CI} , car il contient l'atome $Cro \geq 1$. Après repliement, on obtient le réseau contenu dans la figure 4.11 dont la dynamique est représentée en Figure 4.12.

Dans ce nouveau réseau, nous pouvons regrouper les multiplexes m_{Cro1} et m_{Cro2} dans un seul multiplexe, puisqu'il faut que ces deux multiplexes soient ressources de Cro en même temps pour changer la valeur de celui-ci (voir les valeurs des paramètres $K_{Cro, \dots}$).

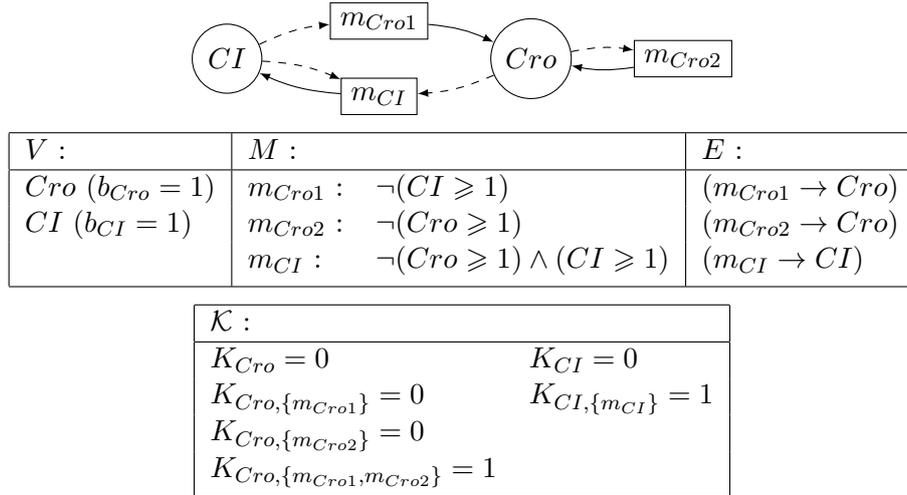


FIGURE 4.11 – Réseau génétique obtenu par suppression du seuil 1 de la variable Cro à partir du réseau de la figure 4.9

Ces deux multiplexes sont remplacés par un nouveau multiplexe m_{Cro} ayant pour formule $\neg(CI \geq 1) \wedge \neg(Cro \geq 1)$. Les paramètres $K_{Cro, \{m_{Cro1}\}}$ et $K_{Cro, \{m_{Cro2}\}}$ disparaissent, le paramètre $K_{Cro, \{m_{Cro1}, m_{Cro2}\}} = 1$ est remplacé par le paramètre $K_{Cro, \{m_{Cro}\}} = 1$, réduisant le nombre de paramètres à 4. Cela permet d'obtenir le réseau de la figure 4.13.

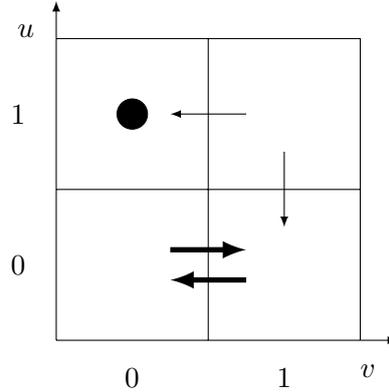


FIGURE 4.12 – Dynamique du réseau de la figure 4.11

4.4 Conclusion

Les réductions successives ont permis d'obtenir, à partir d'un réseau contenant 4 variables et 48 états discrets, un réseau à 2 variables et 4 états tout en conservant les propriétés asymptotiques de la dynamique. Le modèle à 2 variables est donc une abstraction du modèle à 4 variables.

Le nombre de paramètres a également diminué (de 10 à 4), ce qui simplifie grandement le modèle, alors que le formalisme de Naldi augmente largement ce nombre. La différence

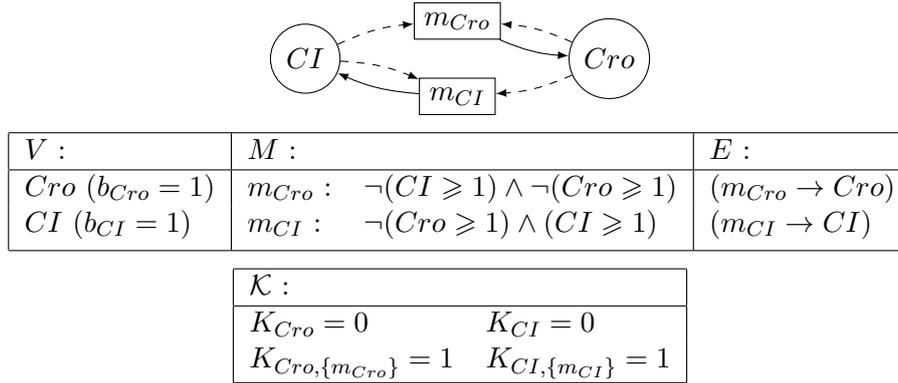


FIGURE 4.13 – Réseau génétique de la figure 4.11 après la fusion des multiplexes m_{Cro1} et m_{Cro2} .

majeure entre les deux approches de réduction réside dans le fait que les informations de la réduction sont stockées dans les formules des multiplexes pour notre méthode, alors que ces informations sont stockées dans les paramètres pour la méthode de Naldi. Dans notre exemple, nous avons simplifié à la volée les formules des multiplexes, car nous connaissons les paramètres. Mais dans la version symbolique, plus un réseau est réduit, plus les formules ont une taille importante. Cependant, ces formules de grande taille restent plus pertinentes qu'un grand nombre de paramètres, car elles explicitent les conditions de régulation prenant en compte les abstractions faites lors des réductions successives.

L'ordre des réductions suivi dans ce chapitre n'est pas le seul possible, et d'autres séquences de réductions ne donneraient pas nécessairement le même réseau final. Dans le cas des suppressions de variables, Adrien Richard a montré que le réseau irréductible dépend de l'ordre de suppression des variables [31]. Comme la notion de réduction est basée sur une intuition de variables très rapides par rapport aux autres, nous pouvons nous demander si ces réductions peuvent représenter des connaissances expérimentales particulières. Par exemple, l'ajout artificiel d'une protéine dans une cellule lors d'une expérimentation peut entraîner une accélération d'une entité cellulaire correspondant à un passage rapide d'un seuil dans le formalisme. Par conséquent, ce choix de l'ordre des réductions pourrait être guidé par l'ensemble des connaissances sur la dynamique du système, en prenant en compte des situations dans lesquelles certaines régulations sont accélérées.

Chapitre 5

Réseaux génétiques hybrides

Le formalisme discret introduit par René Thomas dans les années 1970 est un outil qui permet une vision de la dynamique des réseaux génétiques rapidement appréhendable, car assez simple. De plus, le fait que l'ensemble des états soit fini facilite une identification automatique des paramètres compatibles avec les connaissances biologiques. Cependant, dans beaucoup de réseaux biologiques, la notion de temps est très importante ; or ce formalisme ne la prend pas en compte. Plusieurs formalismes avec délais ont été proposés, mais, soit ces approches sont trop simples pour prendre en compte des phénomènes d'accumulation (voir la fin de la section 5.1), soit leur sophistication les rend difficilement appréhendables.

Dans ce chapitre, nous présentons une approche hybride dans laquelle nous avons essayé de rester le plus proche possible du formalisme de Thomas pour garder sa simplicité tout en prenant en compte les phénomènes d'accumulation.

5.1 Formalismes hybrides pour les réseaux génétiques

5.1.1 Motivations

Beaucoup de problématiques de la biologie ont une nature qualitative dans lesquelles les valeurs *exactes* de concentration, vitesse de synthèse et de dégradation, *etc.* interviennent assez peu dans le processus de compréhension des phénomènes. Des éléments de base qualitatifs tels que la surexpression d'une protéine ou l'activation d'un gène suffisent souvent à établir des causalités explicatives. Dans le but de modéliser ce type de problématiques, les formalismes discrets sont parfaitement adaptés : ils permettent de représenter efficacement les causalités les plus importantes. Ces formalismes de modélisation discrets, en particulier celui de René Thomas, ont pour avantage de nécessiter un nombre limité de paramètres ainsi qu'un nombre d'états finis. Cette caractéristique permet l'utilisation de la logique formelle et du model checking [15, 13] pour l'identification des paramètres des modèles compatibles avec les connaissances biologiques. La possibilité d'effectuer une identification automatique des paramètres facilite grandement l'utilisation de ces formalismes pour la biologie.

Cependant, la notion de temps reste très importante en biologie. En effet, dans des systèmes biologiques tels que le cycle circadien, ou le cycle cellulaire, le temps est souvent un critère important de validation ou d'invalidation d'un modèle : si le cycle circadien n'oscille pas autour de 24 heures, il ne pourra pas assurer correctement son rôle d'anticipation des changements quotidiens de l'environnement. Un modèle formel ne respectant pas cette contrainte serait invalide.

Les formalismes purement discrets ne considèrent le temps que comme une succession d'événements, sans prendre en compte les délais entre ces événements. C'est donc

une modélisation du temps *chronologique*. Comme le montre l'exemple du cycle circadien, cela peut constituer une perte d'information importante dans la modélisation de ces phénomènes biologiques. Ajouter des informations *chronométriques* sur les successions d'événements des modèles discrets, suffirait à affiner de manière satisfaisante les critères de validation ou invalidation du modèle (formalisme hybride).

Les formalismes les plus communément admis pour une modélisation chronométrique sont les formalismes continus, tels que les équations différentielles. La plupart ont le désavantage de ne pas se contenter d'abandonner la discrétisation du temps, ils abandonnent également les espaces d'états discrets et finis. De ce fait, ces formalismes requièrent un nombre important d'informations qui ne sont pas forcément utiles dans la compréhension globale du phénomène.

Dans la Figure 5.1, la partie de gauche représente la dynamique discrète d'un réseau et la partie de droite, une dynamique par équations différentielles compatibles. Dans le cas discret, de l'état (n, m) le système peut évoluer vers les états $(n - 1, m)$ ou $(n, m - 1)$, indépendamment de l'état duquel il vient. Dans le cas continu représenté ici, si le système vient de l'état $(n, m + 1)$, il va obligatoirement aller vers $(n - 1, m)$ et, symétriquement, s'il vient de $(n + 1, m)$, il ira vers $(n, m - 1)$. Ainsi, ajouter une notion temporelle dans un formalisme discret sans prendre en compte le point d'entrée dans un état discret peut générer une incompatibilité avec des comportements de modèles continus. En revanche, la courbure des trajectoires n'est pas nécessairement une information importante. Il est préférable de se focaliser sur la relation entrée-sortie entre les états discrets. Un formalisme hybride qui se contenterait de définir la position de la séparatrice en pointillés montrée dans la figure serait déjà largement utile en biologie.

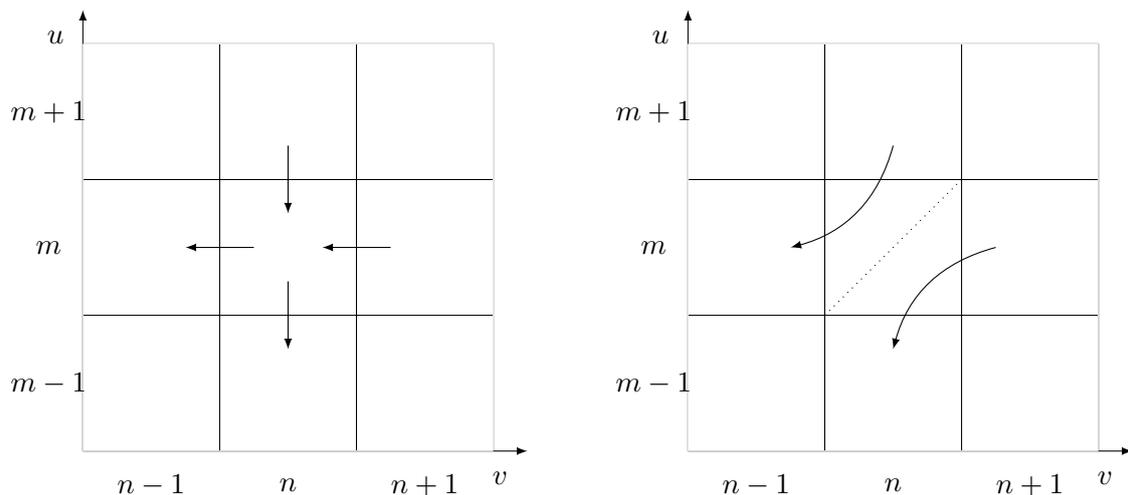


FIGURE 5.1 – Comparaison de la dynamique d'un modèle discret et d'une dynamique compatible dans un formalisme continu. A gauche, la dynamique discrète. A droite, la dynamique continue.

Notre motivation dans ce chapitre est donc de proposer une extension hybride de la

théorie discrète de Thomas respectant ce principe et permettant de représenter des délais. Une préoccupation majeure est de préserver la possibilité d'une identification automatique des paramètres, qu'ils soient entiers ou non. L'idée de base est d'ajouter à l'approche de Thomas, les vitesses d'évolution des variables en fonction de leurs niveaux discrets d'expression.

5.1.2 Principaux formalismes hybrides existants pour les réseaux génétiques

Dans ses travaux, René Thomas s'était déjà penché sur la question de la prise en compte un temps chronométrique dans son formalisme [70, 43, 71]. L'idée est de tenir compte du délai de synthèse d'une protéine ainsi que le délai de dégradation, mais cela n'a jamais été vraiment formalisé, au sens des méthodes formelles de l'informatique.

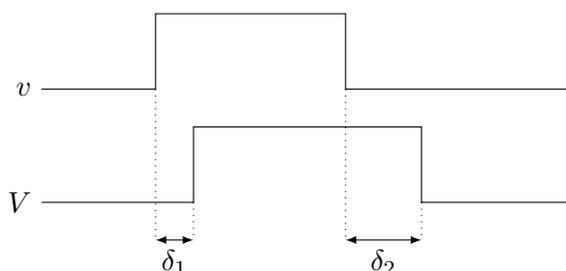


FIGURE 5.2 – Notion de délais présentée par Thomas. δ_1 représente le délai entre l'activation du gène v et le passage du seuil de concentration de la protéine V . δ_2 représente le délai entre l'inhibition de v et le passage du seuil de concentration de V dû à la dégradation de la protéine.

La Figure 5.2 représente la notion de délai proposée par Thomas. L'idée est que l'action d'une "variable" sur une autre n'est pas immédiate. En effet, entre le moment où le facteur de transcription atteint une concentration suffisante pour activer le gène et le moment où le produit du gène passe le seuil d'activation de sa cible, un premier délai δ_1 est nécessaire. De manière similaire, entre le moment où le gène perd ses ressources et le moment où la protéine est dégradée, il se passe un second délai δ_2 .

Le manque de formalisation de cette approche se traduit en partie par le fait que les délais δ_1 et δ_2 devraient dépendre du niveau de concentration discret où se trouve la protéine ainsi que dans la gestion d'un changement de ressources qui arriverait pendant ces délais δ_1 et δ_2 .

Le formalisme des équations linéaires par morceaux [25, 64] peut être considéré comme un formalisme hybride. En effet, les espaces de concentration possibles pour chaque variable sont découpés en intervalles, ce qui fournit donc des "hyperpavés" (états discrets) dans le produit cartésien de ces espaces de concentration possibles. Chaque intervalle de chaque variable est le lieu d'une équation différentielle linéaire, de sorte que chaque état discret

est le lieu d'un système d'équations différentielles linéaires avec autant d'équations que de dimensions (de variables). Chaque équation différentielle est typiquement de la forme :

$$\frac{dv}{dt} = k_0 + \sum_i k_i - \gamma v$$

où les v_i sont les prédécesseurs de v dans le graphe d'interaction ; k_i est une constante strictement positive, si v_i est une ressource de v (au sens habituel de Thomas) et $k_i = 0$ sinon.

La principale faiblesse de cette approche est la difficulté à identifier les paramètres k_i . En effet, chacun des paramètres du système d'équations différentielles présent dans chaque état discret possède la même richesse d'information que dans le cas purement continu. La difficulté d'identification des paramètres est presque la même que dans les équations différentielles standard, à laquelle s'ajoute la difficulté de "recollement" des trajectoires lors des changements d'états discrets (voir Figure 5.3). L'avantage de ces approches hybrides est de garder une bonne précision par rapport aux systèmes d'équations différentielles générales (fonction de Hill typiquement). En l'état actuel des connaissances, une identification automatique des paramètres semble inabordable.

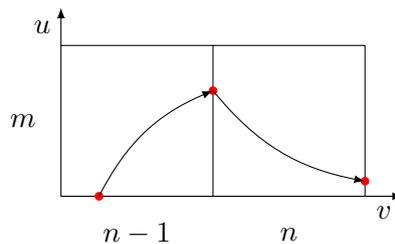


FIGURE 5.3 – Points de "recollement" dans une trajectoire d'un modèle d'équations différentielles linéaires par morceaux

Dans les années 90, Radjeev Alur [5, 4] a proposé un cadre de modélisation hybride sous la forme d'automates temporisés. Bon nombre d'autres équipes ont suivi cette voie [36, 35, 50]. L'avantage des automates temporisés est la décidabilité des formules temporelles. Ceux-ci facilitent grandement l'identification des paramètres. En revanche, toutes les horloges évoluent à la même vitesse dans tous les états discrets. De plus, à l'entrée d'un nouvel état discret, il est uniquement possible de remettre une horloge à 0 ou de conserver sa valeur à la sortie de l'état précédent. Ces deux propriétés rendent les trajectoires non crédibles biologiquement. En effet, lors du passage d'un état discret à un autre, il faut préserver le taux de concentration de chaque protéine, la durée de production ou de dégradation.

À la même époque, Heike Seibert et Alexander Bockmayr ont présenté une adaptation des automates hybrides d'Alur pour le formalisme de René Thomas [62, 61]. Cette approche est fondée sur des produits d'automates, ce qui conduit à surmultiplier le nombre d'états,

mais permet de préserver la décidabilité.

L'idée principale est la suivante. Chaque variable v est associée à une horloge h_v qui évolue linéairement. Pour chaque état discret, il existe 3 comportements possibles en fonction du délai :

- Soit le paramètre $K_{v,\omega}$ (où ω est l'ensemble des ressources de v à l'état courant) est plus grand que la valeur courante de v . Dans ce cas, h_v mesure le temps de synthèse de v .
- Soit le paramètre $K_{v,\omega}$ est plus petit que v , alors h_v mesure le temps de dégradation de v .
- Soit le paramètre $K_{v,\omega}$ est égal à v , alors h_v est arrêté.

Le changement de valeur de v réinitialise l'horloge h_v qui s'arrête ou recommence à mesurer le temps d'évolution en fonction de la valeur de v et du paramètre.

Le principal défaut de toutes les approches présentées jusqu'ici est *l'absence d'accumulation*. En effet, supposons qu'une variable v soit un activateur de u et que v soit soumise à un comportement périodique de telle sorte que v active, puis n'active plus u de manière répétitive. Selon que la vitesse de production de u en présence de v est plus ou moins forte comparée à la vitesse de dégradation de u en l'absence de v , il est possible que u s'accumule sur le long terme (voir Figure 5.4, gauche), ou au contraire se dégrade.

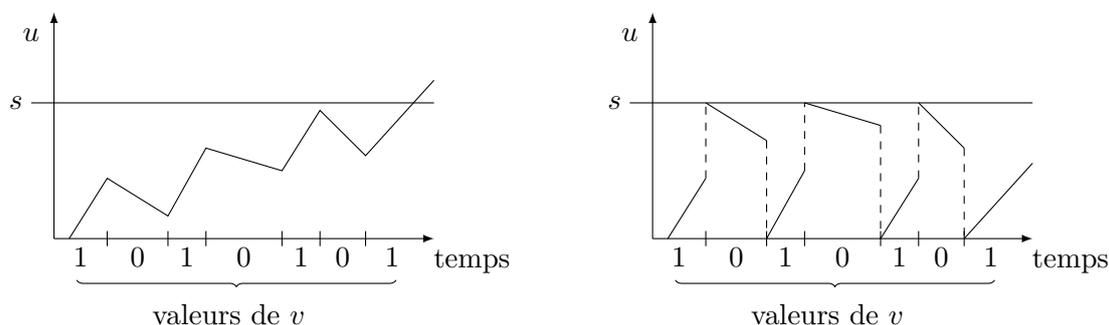


FIGURE 5.4 – La figure de gauche représente une trajectoire prenant en compte l'accumulation de la protéine de u et la figure de droite représente la trajectoire ne prenant pas en compte l'accumulation. On suppose que v , l'activateur de u , s'allume et s'éteint alternativement, donc la concentration de la protéine de u monte et descend alternativement. Ne pas prendre en compte l'accumulation empêche le passage du seuil de u .

Il est clair qu'avec une faible dégradation de u , après un nombre suffisant d'alternances de l'action de v sur u , u finira par passer son seuil supérieur, et inversement, avec un fort taux de dégradation, son seuil inférieur. Tous les formalismes décrits précédemment sont dans l'incapacité de modéliser ces passages de seuils, parce que les horloges sont réinitialisées à chaque changement de signe de la dérivée de u . Par conséquent, u reste dans le même niveau discret (voir Figure 5.4, droite).

Ce problème récurrent de non pris en compte de l'accumulation a été résolu grâce à une collaboration entre l'équipe à laquelle j'appartiens et celle d'Olivier Roux, au

prix de la perte la décidabilité [22, 2]. Dans ces théories, des « règles de trois » sont systématiquement appliquées lors des changements d'états discrets pour préserver l'échelle des taux de concentration plutôt que l'échelle des temps. Les vérifications concernent principalement l'identification des régions invariantes ainsi que sur l'extraction de contraintes sur les délais à partir de la connaissance des durées d'un ensemble de chemins. Ce cadre de modélisation peut être couplé à HyTech [3], ou des model checkers d'automates hybrides. Malgré l'indécidabilité, cela outille utilement l'identification des paramètres. La principale faiblesse de cette approche *par règle de trois* est la lourdeur du formalisme et un certain nombre de redondances qu'il induit. Ce cadre de modélisation est basé sur les travaux de Thomas, auxquels ont été ajoutés des délais spécifiques à chacun des états discrets. Ces délais sont représentés par un changement de la taille des états (plus le délai est important, plus l'état sera grand). Contrairement aux formalismes présentés plus haut, l'état précédent est pris en compte dans l'état courant, afin de déterminer le suivant. Cependant, l'ajout du délai en plus des paramètres de Thomas augmente considérablement le nombre de paramètres du modèle, ce qui nous éloigne de la simplicité du formalisme de Thomas.

Notre approche « avec célérités », s'inspire en partie de celle *par règles de trois* [22, 2] et s'attache à réduire la lourdeur du formalisme en réduisant le nombre de paramètres à identifier à son strict minimum, préservant ainsi l'élégance de l'approche d'origine de René Thomas.

5.2 Réseaux génétiques à célérité

Les travaux présentés à partir de cette section constituent la seconde partie de mon travail de doctorat. Nous cherchons à obtenir un cadre de modélisation hybride qui, d'une part, gère les points d'entrée dans les états discrets de sorte que les phénomènes d'accumulation soient pris en compte, et d'autre part, limite le nombre de paramètres du modèle, typiquement dans l'état d'esprit de la théorie de Thomas.

L'idée est de conserver l'espace des états discrets de la théorie de Thomas, découpé de manière cartésienne et unitaire, et d'intégrer les délais de traversée des états *via* des « célérités ». Comme on le verra plus tard, ces informations suffisent pour trouver la valeur focale d'une variable en fonction de celles qui agissent sur elle et apportent donc la direction d'évolution de la variable. Une célérité est une vitesse abstraite : le passage du système d'une face d'un état discret à sa face opposée nécessite un délai, autrement dit, l'inverse de ce délai est la célérité.

Pour donner un exemple, supposons que, dans un état discret donné, l'on sache grâce aux connaissances biologiques qu'une variable v nécessite 5 heures entre le moment où elle entre dans un état n et le moment où elle en sort, toutes choses égales par ailleurs. Si ses ressources font qu'elle est croissante (resp. décroissante), alors la vitesse de cette variable

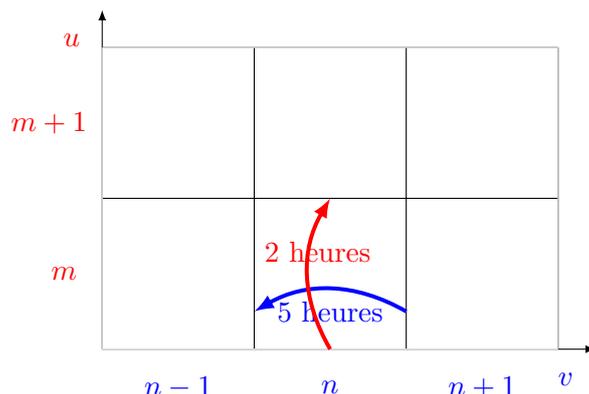


FIGURE 5.5 – Illustration des délais entre deux faces opposées d'un état discret

sera intuitivement de $\frac{1}{5}$ (resp. $-\frac{1}{5}$). Avec les célérités, nous avons l'information temporelle (la valeur de la vitesse et son signe), mais aussi, comme nous le verrons plus loin, les paramètres de Thomas. De plus, supposons que dans le même état discret, une variable u ne nécessite que 2 heures pour passer d'une face à l'autre, comme dans la figure 5.5. Dès lors, v ne pourra jamais traverser entièrement cet état puisque u sera plus rapide : en ce sens, le délai de 5 heures est un délai abstrait.

Comme on le voit sur cet exemple, la célérité $-\frac{1}{5}$ se calcule en considérant que traverser un état discret, est équivalent à parcourir une longueur de 1. Comme nous l'avons vu dans le chapitre 1, le découpage en intervalles de la plage de concentration possibles d'une variable, est fondé sur les seuils d'interaction entre gènes, et n'est par conséquent presque jamais equiréparti. Il est donc possible qu'une vitesse de $\frac{1}{5}$ induise une variation de concentration bien plus rapide qu'une vitesse de 1 dans un autre intervalle : il suffit pour cela que la plage de concentration que représente le premier intervalle, soit au moins 5 fois plus grande que celle que représente le deuxième. C'est pour cette raison, et pour limiter les confusions qu'il est préférable d'adopter le terme de « célérité » plutôt que « vitesse ». Une *célérité* est, dans notre contexte, un concept abstrait, alors qu'une « vitesse » pourrait porter une signification quantitative liée à la vitesse de production/dégradation du produit d'un gène.

Dans le but de conserver la simplicité du formalisme de René Thomas, la définition des réseaux de régulations génétiques hybrides présentée dans cette section a une structure très proche de celle présentée dans le chapitre 1. Le lecteur retrouvera donc de nombreux concepts et de nombreuses intuitions déjà détaillées dans ce dernier. Nous allons par conséquent donner dès maintenant les définitions formelles et essentiellement commenter les aspects hybrides du formalisme.

Du chapitre 1, nous conservons en particulier la définition du langage des multiplexes :

Étant donné un ensemble V de variables, l'ensemble des formules de multiplexes \mathcal{L}_V

est défini par :

- Si $v \in V$ et n est un entier tel que $1 \leq n \leq b_v$, où b_v est la borne de v , alors $v \geq n$ est un atome de \mathcal{L}_V .
- Si φ et ψ sont deux formules de \mathcal{L}_V , alors $\neg\varphi$, $(\varphi \vee \psi)$, $(\varphi \wedge \psi)$ et $(\varphi \Rightarrow \psi)$ appartiennent aussi à \mathcal{L}_V .

De plus, la notation suivante sera utile :

$$\text{— La fonction } \textit{sgn} : \mathbb{R} \rightarrow \{-1, 0, 1\} \text{ est définie par : } \textit{sgn}(x) = \begin{cases} -1 & \text{si } x < 0 \\ 0 & \text{si } x = 0 \\ 1 & \text{si } x > 0 \end{cases}$$

Sur la base des idées brièvement présentées précédemment, on comprend que, comme les points focaux pourront être déduits des célérités, les paramètres de Thomas deviennent redondants face à la connaissance de ces célérités. Ainsi, dans la définition des réseaux de régulations génétiques hybrides (abrévés en RH pour « Réseaux Hybrides »), on remplace la famille de paramètres discrets \mathcal{K} par une famille de célérités \mathcal{C} .

Définition 13. [*Réseau de régulation génétique hybride*] *Un réseau de régulation génétique hybride est un quadruplet $R = (V, M, E, \mathcal{C})$ où :*

- V est un ensemble fini de variables dans lequel chaque $v \in V$ est muni d'une borne $b_v \in \mathbb{N}^*$.
 - M est un ensemble fini de multiplexes dans lequel chaque $m \in M$ est muni d'une formule φ_m appartenant au langage \mathcal{L}_V .
 - E est un ensemble d'arcs de la forme $(m \rightarrow v) \in M \times V$.
 - $\mathcal{C} = \{C_{v,\omega,n}\}$ est une famille de nombres réels indexés par l'ensemble des triplets (v, ω, n) où v , ω et n vérifient les 3 conditions suivantes :
 1. $v \in V$
 2. ω est un sous-ensemble quelconque de $R^-(v)$ où $R^-(v) = \{m \mid (m \rightarrow v) \in E\}$.
Autrement dit, ω est un sous-ensemble de prédécesseurs de v .
 3. n est un entier tel que $0 \leq n \leq b_v$
- $C_{v,\omega,n}$ est la célérité de v au niveau n pour l'ensemble de ressources ω .

De plus, la famille \mathcal{C} est soumise aux contraintes suivantes pour tout $v \in V$ et tout $\omega \subset R^-(v)$:

- Soit, pour tout n compris entre 0 et b_v inclus, les célérités $C_{v,\omega,n}$ ont le même signe non nul
- Soit il existe un n_0 compris entre 0 et b_v inclus vérifiant les 3 conditions suivantes :
 - $C_{v,\omega,n_0} = 0$
 - Si $0 \leq n < n_0$ alors $\textit{sng}(C_{v,\omega,n}) = 1$
 - Si $n_0 < n \leq b_v$ alors $\textit{sng}(C_{v,\omega,n}) = -1$

Il est à noter que la famille des célérités requiert un index supplémentaire par rapport aux index des paramètres discrets de Thomas. Cet entier n est le niveau d'expression de

la variable v . Dans le cadre discret, si l'ensemble des ressources ω de la variable v est constant, la valeur focale vers laquelle la variable v est attirée est elle-même constante quelque soit n : l'entier $K_{v,\omega}$.

- Dans la seconde contrainte sur la famille des célérités, cet entier n_0 n'est autre que $K_{v,\omega}$. La contrainte sur le signe des célérités dit simplement que v est attiré par n_0 (voir Figure 5.6).

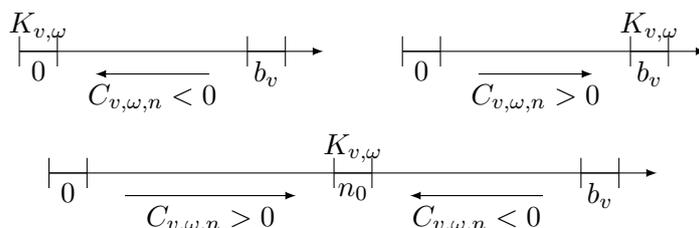


FIGURE 5.6 – Relation entre les paramètres discrets de Thomas et le signe des célérités

- Supposons par exemple que $n_0 = 3$. Il est évident que si la variable v part de 0 pour atteindre 3, le temps de traversée de l'intervalle 1 n'a aucune raison d'être le même que le temps de traversée de l'intervalle 2. Cette différence existe même à ressources constantes, ne serait-ce que parce que le découpage de l'espace des concentrations continues n'est pas équiréparti. Cela implique qu'il faille distinguer $C_{v,\omega,1}$ et $C_{v,\omega,2}$. Ceci justifie donc pleinement l'index supplémentaire n .

Une remarque importante : dans cette définition, nous avons étendu la théorie de Thomas en autorisant le fait que toutes les célérités soient de même signe, ce qui correspondrait intuitivement à accepter des paramètres discrets $K_{v,\omega}$ tel $K_{v,\omega} = -1$ ou $K_{v,\omega} = b_v + 1$. Comme on le verra plus tard, ce qui a motivé cette extension est la réduction de la taille de l'espace des états grâce à des « glissements » sur les bords. Cette extension de la théorie de Thomas a pour conséquence qu'il n'est pas possible d'appliquer aveuglément un certain nombre de théorèmes établis dans le cadre purement discret.

L'apport d'observations biologiques est primordial dans l'identification des célérités. Cependant, ces observations ne permettent que très rarement de déterminer *directement* une célérité donnée. Dans beaucoup de cas, les délais obtenus expérimentalement correspondent à des successions de plusieurs états adjacents le long d'une trajectoire. L'accumulation des observations fournit donc un système d'équations sur des sommes de délais dont le total doit être égal à la durée de chaque trajectoire observée. En pratique, lorsqu'il y a plus de délais à identifier que le système d'équations ne le permet, le modélisateur applique souvent une équirépartition des délais sur tous les états correspondants.

Rappelons également que, même si les vitesses réelles de variation des concentrations étaient égales dans tous les intervalles de Thomas, elles ne le seraient plus après normalisation sous forme de célérités. Inversement, deux célérités égales ne représentent donc pas nécessairement deux vitesses égales de variation de concentration. La célérité d'une

$V :$	$M :$	$E :$	$\mathcal{C} :$
$v (b_v = 1)$	$m_u : (v \geq 1)$	$(m_u \rightarrow v)$	$C_{v,\{\},0}$ $C_{u,\{\},0}$
$u (b_u = 1)$	$m_v : \neg(u \geq 1)$	$(m_v \rightarrow u)$	$C_{v,\{\},1}$ $C_{u,\{\},1}$
			$C_{v,\{m_u\},0}$ $C_{u,\{m_v\},0}$
			$C_{v,\{m_u\},1}$ $C_{u,\{m_v\},1}$

TABLE 5.1 – Contenu des ensembles du réseau de la figure 5.7.

variable lente sur un intervalle très court pourra être plus élevée que la célérité d'une variable rapide sur un intervalle très grand.

Dans le même ordre d'idée, nous savons que plus une variable se rapproche de sa valeur focale, plus sa vitesse diminue, mais cela n'implique pas pour autant que sa célérité diminue.

Exemple

Dans la figure 5.7, le réseau est composé de 2 variables, u et v qui forment un cycle négatif dans lequel u inhibe v qui active u . Il y a 2 multiplexes m_v et m_u qui représentent respectivement l'action de v et de u sur leur cible. Les arcs de la définition sont représentés par les flèches pleines, alors que les flèches en pointillés ne font que rappeler la participation des variables dans la formule du multiplexe.

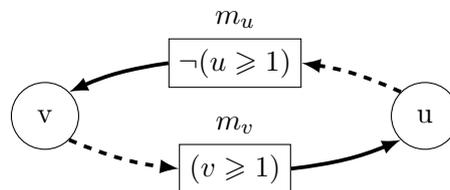


FIGURE 5.7 – Graphe d'interaction d'un réseau génétique hybride à 2 variables

Le tableau 5.1 contient la description des ensembles V , M , E et \mathcal{C} du réseau représenté par la figure 5.7. Le nombre de paramètres est plus important que dans le formalisme de Thomas : 8 au lieu de 4. Chacun des paramètres de Thomas est multiplié par le nombre de valeurs que peut prendre la variable (ici deux, car les variables sont booléennes).

Un des avantages des multiplexes est la représentation d'un ensemble d'interactions par une formule logique afin de diminuer le nombre de paramètres, qui dans notre cas, peut augmenter très vite. Le fait de pouvoir manipuler des formules logiques facilite grandement les calculs formels, surtout lorsqu'un multiplexe contient plusieurs littéraux.

À l'opposé, lorsqu'un même processus contrôle plusieurs variables, un multiplexe avoir plusieurs successeurs. C'est par exemple le cas pour les opérons où un groupe de gènes adjacents est transcrit à partir d'un unique promoteur. Tous les gènes de cet opéron sont

donc régulés simultanément.

Une remarque importante : dans le cadre hybride, il faut ajuster la dynamique plus finement que dans le formalisme de Thomas. Or si plusieurs actions sont regroupées au sein d'un même multiplexe, il n'y aura qu'une seule célérité associée à tous les états qui satisfont (resp. ne satisfont pas) le multiplexe. Il est parfois nécessaire d'avoir à « découper » un multiplexe pour identifier plus fidèlement les célérités.

Imaginons par exemple deux variables a et b telles que la présence de l'une d'elles soit suffisante pour faire passer son seuil à une variable booléenne c (voir Figure 5.8 gauche). Lorsque la formule du multiplexe m ($a \geq 1 \vee b \geq 1$) est vraie, la variable c va augmenter lorsqu'elle vaut 0, c'est-à-dire $C_{c,m,0} > 0$. Il se peut cependant que la présence de a et de b conjointement accélère la production de c , de sorte qu'il est alors judicieux de séparer m en deux multiplexes m_a et m_b (Figure 5.8 droite). Ceci permet d'avoir trois vitesses positives différentes $C_{c,\{m_a\},0}$, $C_{c,\{m_b\},0}$ et $C_{c,\{m_a,m_b\},0}$ au lieu de la seule et unique vitesse positive $C_{c,m,0}$.

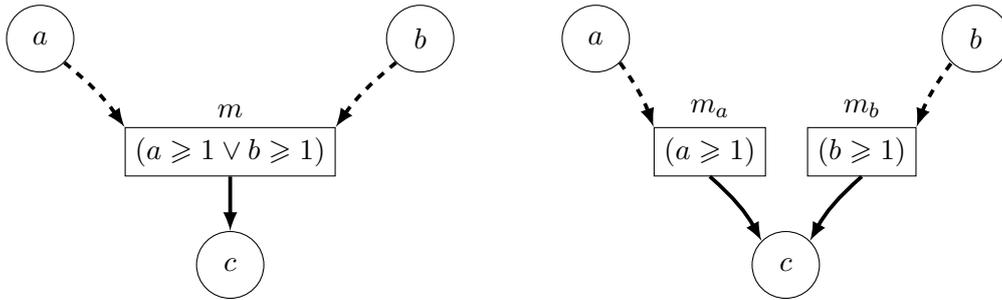


FIGURE 5.8 – Découpe d'un multiplexe lorsque les célérités dépendent de la configuration

5.3 États hybrides, ressources et quelques propriétés

Pour définir la notion d'état dans ce formalisme hybride, il est nécessaire, pour chaque variable v , de connaître dans quel état discret elle se trouve et nous avons également besoin de connaître son niveau de progression au sein de cet état. Nous proposons donc la définition suivante.

Définition 14. [*États d'un réseau*] Soit $R = (V, M, E, C)$ un RH. Un état hybride de R est un couple $h = (\eta, \pi)$ où :

- η est une fonction de V dans \mathbb{N} telle que pour tout $v \in V$, $0 \leq \eta(v) \leq b_v$;
 η est appelé l'état discret de h .
- π est une fonction de V dans l'intervalle $[0, 1]$ des nombres réels ;
 π est appelé la partie fractionnaire de h .

On note H l'ensemble des états hybrides de R .

Quand il n'y a pas d'ambiguïté, nous utiliserons η et π sans mentionner explicitement h .

Chaque état hybride est donc constitué de 2 parties : la partie discrète η correspond aux états du formalisme de Thomas, et la partie fractionnaire π donne la position du système à l'intérieur de l'hypercube qu'est cet état discret.

L'intervalle fermé $[0, 1]$ de la partie fractionnaire permet d'obtenir des « doubles seuils » pour chacun des seuils de l'espace des états : $(\eta(v) = n, \pi(v) = 1)$ et $(\eta(v) = n + 1, \pi(v) = 0)$ sont deux états distincts de la variable v , alors que dans la théorie de Thomas, il n'y a qu'un seul état singulier entre n et $n + 1$ (le seuil). Nous verrons l'intérêt de cette distinction plus tard.

Tout comme dans le cadre discret utilisé jusqu'à maintenant dans cette thèse, les ressources sont des multiplexes (et non des variables).

Définition 15. [Ressources] Soit $R = (V, M, E, \mathcal{C})$ un RH, φ une formule de multiplexe et $h = (\eta, \pi)$ un état hybride de R . La relation de satisfaction $h \models \varphi$ est inductivement définie par :

- Si φ est un atome de la forme $v \geq n$, alors $h \models \varphi$ si et seulement si $\eta(v) \geq n$
- Si φ est de la forme $\neg\psi$, alors $h \models \varphi$ si et seulement si $h \not\models \psi$
- Si φ est de la forme $\psi_1 \vee \psi_2$, alors $h \models \varphi$ si et seulement si $h \models \psi_1$ ou $h \models \psi_2$
- Si φ est de la forme $\psi_1 \wedge \psi_2$, alors $h \models \varphi$ si et seulement si $h \models \neg(\neg\psi_1 \vee \neg\psi_2)$
- Si φ est de la forme $\psi_1 \Rightarrow \psi_2$, alors $h \models \varphi$ si et seulement si $h \models (\psi_2 \vee \neg\psi_1)$.

L'ensemble des ressources d'une variable v pour un état h est défini par : $\rho(h, v) = \{m \in R^-(v) \mid h \models \varphi_m\}$, c'est-à-dire l'ensemble des multiplexes prédécesseurs de v dont la formule est satisfaite.

Dans la définition précédente, l'ensemble des ressources est uniquement dépendant de l'état discret. Par conséquent, pour $v \in V$ et $h = (\eta, \pi)$, l'ensemble des ressources $\rho(h, v)$ sera noté $\rho(\eta, v)$ par abus de langage.

Rappelons que dans le cadre discret, certains paramètres $K_{v,\omega}$ sont inutiles, car ils ne s'appliquent dans aucun état : il peut en effet arriver que la formule

$$\Phi_\omega^v \equiv \left(\bigwedge_{m \in \omega} \varphi_m \right) \wedge \left(\bigwedge_{m \in R^-(v) \setminus \omega} \neg\varphi_m \right)$$

, déjà mentionnée dans le chapitre 2, soit non satisfiable. *A fortiori*, certaines célérités $C_{v,\omega,n}$ sont inutiles puisque $(\Phi_\omega^v \wedge (v = n))$ peut-être non satisfiable. De tels cas sont même plus fréquents que dans la théorie discrète, puisque l'ajout du niveau de la variable associe chaque paramètre à un sous-ensemble d'états discrets plus petits. Par exemple, si il y a une autorégulation positive d'une variable v au seuil n , alors, dans tout état tel que $\eta(v) < n$, ω ne peut pas contenir v . Les célérités $C_{v,\omega,k}$ telles que $k < n$ et ω contient v ne seront jamais utilisées.

Dans notre cadre de modélisation hybride, il est possible d'adapter les contraintes de

Snoussi. Pour rappel, dans le cadre discret, ces contraintes sont satisfaites quand l'ajout d'une ressource ne fait pas diminuer la valeur des paramètres. Dans notre cadre hybride, les contraintes de Snoussi s'énoncent

$$\forall v \in V, \forall \omega \subset R^-(v), \forall \omega' \subset R^-(v), \forall n \in [0 \dots b_v], (\omega \subset \omega' \Rightarrow C_{v,\omega,n} \leq C_{v,\omega',n})$$

Comme dans le cadre discret, ces contraintes peuvent aider à identifier des célérités à partir de connaissances biologiques. Elles viennent alors s'ajouter aux conditions intrinsèques de continuité des célérités qui imposent qu'à ressources fixées, le signe de la célérité doit être décroissant avec n , et ne peut pas passer directement de strictement positif à strictement négatif. En revanche, comme nous l'avons déjà dit, cette continuité ne concerne que le signe des célérités et non leur valeur, notamment en raison de la normalisation des intervalles de concentration.

5.4 Définition de la dynamique

La dynamique d'un réseau hybride se décomposera en une suite de deux types de transitions. En effet, une trajectoire devra successivement traverser un état discret, puis traverser un seuil pour changer d'état discret. Dans cette section, nous allons définir les caractéristiques de chacune de ces transitions, ainsi que les cas particuliers de la dynamique lorsqu'il n'est pas possible de traverser un seuil rencontré. Le dédoublement des seuils déjà mentionné jouera alors un rôle crucial.

5.4.1 Espace des transitions hybrides

Par rapport à un formalisme continu, l'avantage d'une approche hybride avec célérités constantes dans chaque état discret réside dans la simplicité du calcul des délais qui s'appuie sur des propriétés géométriques élémentaires.

Définition 16. [Délai de contact] Soit un RH $R = (V, M, E, C)$, v une variable de V et $h = (\eta, \pi)$ un état hybride. On note $\delta_h(v)$ le délai de contact de v dans h pour atteindre le bord de l'état discret. Plus précisément, δ_h est la fonction de V dans $\mathbb{R}^+ \cup \{+\infty\}$ définie par :

- Si $C_{v,\rho(\eta,v),\eta(v)} = 0$ alors $\delta_h(v) = +\infty$
- Si $C_{v,\rho(\eta,v),\eta(v)} > 0$ alors $\delta_h(v) = \frac{1-\pi(v)}{C_{v,\rho(\eta,v),\eta(v)}}$
- Si $C_{v,\rho(\eta,v),\eta(v)} < 0$ alors $\delta_h(v) = \frac{\pi(v)}{|C_{v,\rho(\eta,v),\eta(v)}|}$

Comme nous avons vu dans le Chapitre 2, dans le formalisme de Thomas, le comportement des autorégulations négatives est représenté par deux transitions opposées sur le seuil de l'autorégulation. Il est donc possible de le voir comme un cycle de longueur 2.

Dans le cadre hybride, ainsi que dans tous les cadres de modélisation continue, cette trajectoire cyclique ne peut pas exister, puisque à chaque état hybride d'un état discret est associé qu'une unique valeur de célérité. Or, pour avoir un aller-retour entre les deux états, il faut changer de signe à l'intérieur d'un état discret. Par conséquent, une trajectoire ne peut pas traverser dans les deux sens un seuil, même s'il y a une autorégulation négative.

De manière similaire, dans les états adjacents aux bornes de l'espace des états, peuvent pointer vers l'extérieur de ces états (Définition 13). Il faut donc empêcher les trajectoires de sortir de l'espace des états.

Définition 17. [*Murs noirs et murs extérieurs*] Soit $R = (V, M, E, C)$ un RH, $v \in V$ une variable, $h = (\eta, \pi)$ et $h' = (\eta', \pi')$ deux états hybrides tels que η et η' sont des états voisins par rapport à v , tels que $\eta'(v) = \eta(v) + \text{sgn}(C_{v,\rho(v,\eta),\eta(v)})$ et $\eta'(u) = \eta(u)$ pour tout $u \neq v$.

1. v est dit sur un mur noir si les deux conditions suivantes sont satisfaites :
 - $\delta_h(v) = 0$
 - $\text{sgn}(C_{v,\rho(v,\eta),\eta(v)}) \times \text{sgn}(C_{v,\rho(v,\eta'),\eta'(v)}) = -1$
2. v est dit sur un mur extérieur si les deux conditions suivantes sont satisfaites :
 - $\delta_h(v) = 0$
 - Au moins une des conditions suivantes est respectée :
 - $C_{v,\rho(v,\eta),\eta(v)} < 0$ et $\eta(v) = 0$
 - $C_{v,\rho(v,\eta),\eta(v)} > 0$ et $\eta(v) = b_v$.

On note $\text{slide}(h)$ l'ensemble des variables qui sont soit sur un mur noir, soit sur un mur extérieur.

Intuitivement, les conditions $\delta_h(v) = 0$ signifient que la trajectoire a atteint un seuil, et le produit des signes négatifs signifie que cette trajectoire « bute » sur une vitesse contraire dans l'état discret voisin. Remarquons qu'un produit des signes négatifs pourrait également dénoter un « mur blanc » (les trajectoires s'éloignent du seuil), mais dans ce cas $\delta_h(v)$ ne pourrait pas être égal à 0. Il s'agit donc bien d'un mur noir vers lequel les deux vitesses vont converger. La situation sur les murs extérieurs est tout à fait similaire.

Pour résumer, l'ensemble slide est l'ensemble de toutes les variables bloquées par la dynamique du système et qui sont donc en attente d'un changement d'état discret. Le nom « slide » a été choisi, car les variables qui ne sont pas dans cet ensemble continuent d'évoluer, ce qui donne un effet de glissement à la trajectoire. Quand la trajectoire atteint le mur, elle change de direction et longe celui-ci. Ce comportement approxime des trajectoires obtenues dans un modèle continu, à l'approche d'une asymptote.

Les glissements sur les murs extérieurs ont une signification biologique. La sémantique d'un glissement sur la borne inférieure de l'espace des états est la dégradation presque

totale de la protéine. À l'inverse, celle d'un glissement sur la borne supérieure correspond à la saturation de cette protéine dans le compartiment cellulaire. Du point de vue expérimental, ces affirmations sont aisément observables et constituent donc des informations qui permettront d'affiner les modèles.

Définition 18. [*Premières variables*] Soit $R = (V, M, E, C)$ un RH et $h = (\eta, \pi)$ un état hybride. L'ensemble des premières variables est défini par :

$$first(h) = \{v \in V \setminus slide(h) \mid \delta_h(v) < +\infty \text{ et } \forall u \in V \setminus slide(h), \delta_h(u) \geq \delta_h(v)\}$$

L'ensemble $first(h)$ contient les variables qui atteindront leur prochain seuil en premier. Dans le cas où la trajectoire arrive dans un coin d'un état discret (si plusieurs variables atteignent leur seuil simultanément), l'ensemble pourra contenir plusieurs éléments. Cependant, le volume de l'ensemble des états hybrides h tel que $|first(h)| \neq 1$ est de mesure nulle.

Définition 19. [*Espace des états hybrides*] Soit $R = (V, M, E, C)$ un RH, on note $\mathcal{R} = (H, T)$ l'espace des états hybrides de R où H est l'ensemble des états hybrides et T est l'ensemble des transitions défini comme suit :

il existe une transition de (η, π) vers (η', π') si et seulement s'il existe une variable $v \in first(h)$ telle que :

1. Soit $\delta_h(v) \neq 0$ et
 - (i) $\eta' = \eta$
 - (ii) $\pi'(w) = \frac{1+sgn(C_{w,\rho(w,\eta),\eta(w)})}{2}$ pour tout $w \in first(h)$ (i.e. 0 si $C_{w,\rho(w,\eta),\eta(w)} < 0$ et 1 si $C_{w,\rho(w,\eta),\eta(w)} > 0$).
 - (iii) pour toute variable $u \in first(h)$ telle que $u \neq v$ si $u \in slide(h)$ alors $\pi'(u) = \pi(u)$ et sinon $\pi'(u) = \pi(u) + \delta_h(v) \times C_{u,\rho(u,\eta),\eta(u)}$.
2. Soit $\delta_h(v) = 0$ et
 - (i) $\eta'(v) = \eta(v) + sgn(C_{v,\rho(v,\eta),\eta(v)})$
 - (ii) $\pi'(v) = \frac{1-sgn(C_{v,\rho(v,\eta),\eta(v)})}{2}$ (i.e. 0 si $C_{v,\rho(v,\eta),\eta(v)} > 0$ et 1 si $C_{v,\rho(v,\eta),\eta(v)} < 0$).
 - (iii) Pour toute variable $u \neq v$, $\eta'(u) = \eta(u)$ et $\pi'(u) = \pi(u)$

Pour chacune des trajectoires, les variables qui génèrent les prochaines transitions sont celles de l'ensemble $first(h)$: celles qui atteignent leur seuil en premier. Ce sont elles qui imposent le délai avant la prochaine transition et les autres variables évoluent en fonction de ce délai.

L'espace des états hybrides porte 2 types de transitions :

- Les transitions continues (cas 1 de la définition) correspondent à la traversée d'un état discret. Il n'y a pas de changement d'état discret (item 1.i de la définition). Par contre, la partie fractionnaire va évoluer. Les variables de $first(h)$ deviendront 0 ou 1 si elles ont respectivement une célérité négative ou positive. Les autres variables,

si elles ne sont pas sur un mur noir ou un mur extérieur, vont simplement évoluer selon leur célérité, le temps que celles de $first(h)$ atteignent leur seuil.

- *Les transitions discrètes* (cas 2 de la définition) correspondent à un changement d'état discret. Les variables de $first(h)$ changent de valeur discrète (+1 si la célérité est positive et -1 si elle est négative) et de valeur fractionnaire (0 si elle est positive et 1 si elle est négative). Par contre, les autres variables ne changent pas. En effet, comme cette transition ne fait que traverser un seuil, elle se fait instantanément.

Une trajectoire est alors constituée d'une succession alternée de ces deux types de transitions.

Comme déjà mentionné, l'intervalle fermé $[0, 1]$ des parties fractionnaires permet de distinguer le seuil entre la valeur $n-1$ et la valeur n d'une variable v , selon que la trajectoire hybride arrive depuis l'état discret $n-1$ ou depuis l'état discret n . Lors des glissements, cela permet à la trajectoire de dépendre de son état discret d'origine. La figure 5.9 gauche illustre l'effet du double mur. Dans l'état où $v = n-1$, quand v atteint le seuil entre $n-1$ et n , c'est la célérité de u dans l'état où $v = n-1$ qui est prise en compte pour le glissement. Comme celle-ci est positive, la trajectoire monte. De la même façon, dans l'état où $v = n$, quand v atteint le seuil entre $n-1$ et n , c'est la célérité de u dans l'état courant qui est prise en compte. Comme elle est négative, la trajectoire descend. Ce phénomène permet d'approximer avec des vitesses constantes des phénomènes continus avec asymptote où les vitesses sont opposées autour de celle-ci, comme illustré dans la figure 5.9 droite.

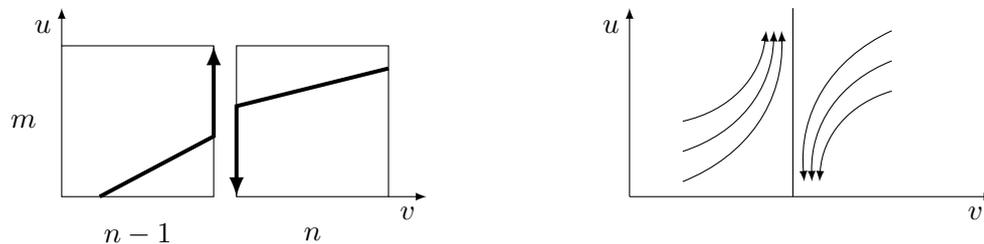


FIGURE 5.9 – Trajectoires autour d'un mur noir. Le glissement sur le seuil dépend de l'état discret d'origine de la trajectoire. À gauche, la trajectoire hybride et à droite une trajectoire dérivable correspondante.

5.4.2 Le cas des « hypercoins »

Il s'agit d'états hybrides dans lesquels toutes les variables qui ne sont pas sur un mur noir ont une célérité nulle. Par exemple, dans la figure 5.10 gauche, l'état en rouge ($\eta(v) = 1$, $\eta(u) = 1$, $\pi(v) = 1$ et $\pi(u) = 0,5$) a une célérité nulle en v et strictement positive en u , mais u est sur un mur noir. Cet état est un hypercoin parce qu'il est bloqué (il ne fait l'objet d'aucune transition). La partie droite de la figure 5.10 représente une situation similaire en 3 dimensions en ne représentant que l'état discret concerné.

La définition 19 fait le choix d'un comportement aussi simple que possible dans les

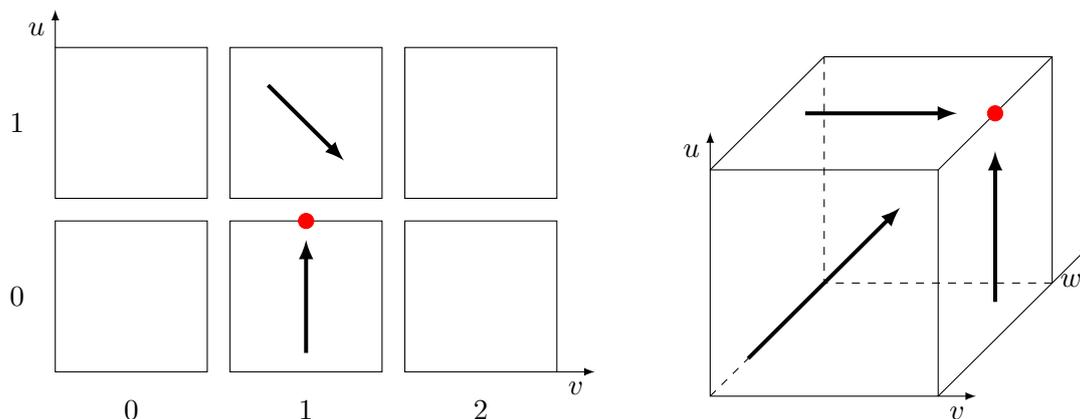


FIGURE 5.10 – Dynamique d’un réseau dans le cas d’hypercoins. À gauche : en 2 dimensions, dans l’état $(1,0)$, la variable v a une vitesse nulle alors que le point rouge est situé sur un mur noir en u . À droite : en 3 dimensions, la vitesse en w est nulle, c’est-à-dire que les 3 flèches noires sont dans un même plan généré par les vecteurs u et v , et l’on suppose que les cubes autour de l’arête du point rouge imposent une « arête noire » en u et v .

hypercoins, pour des raisons de simplicité. Ces choix ont des implications qui peuvent être en désaccord avec des trajectoires dérivables. En effet, s’il est tout à fait raisonnable d’avoir un hypercoin stable dans la figure 5.11 gauche (représenté par le point rouge), nos définitions rendent stable l’état rouge de la figure 5.11 droite, alors que les trajectoires dérivables plaident plutôt pour une « fuite en avant » des états de l’état discret $(n-1, m-1)$ vers l’état discret (n, m) (de l’état hybride en rouge à l’état hybride en bleu).

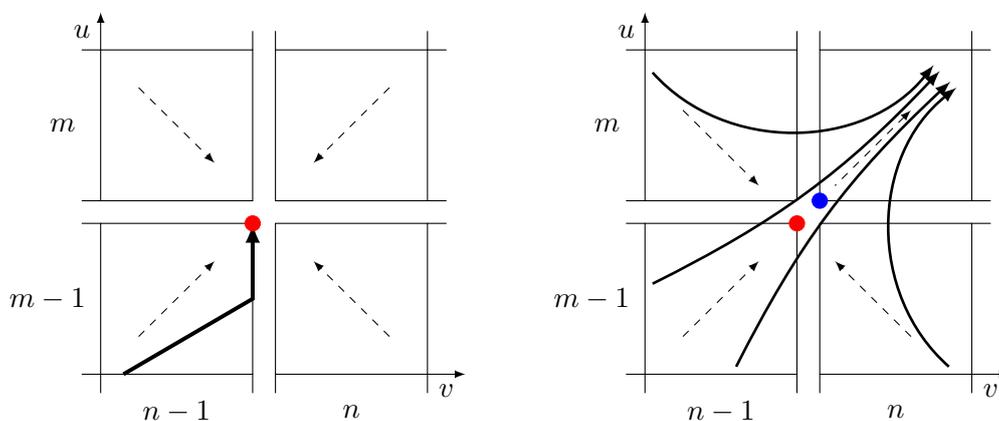


FIGURE 5.11 – Dynamiques issues du graphe de la figure 5.12 sans (à gauche) et avec (à droite) les multiplexes gris

Pour adapter la définition 19 de telle sorte que l’on admette une transition du point rouge vers le point bleu, il faudrait autoriser des changements d’états discrets *synchrones* lorsqu’on rencontre de tels cas particuliers (u et v changent d’états discrets en même temps). En effet, la solution « simple » qui consisterait à autoriser le changement d’état discret dans un mur noir conduirait à des chemins admettant un nombre infini de transitions discrètes en un temps 0. Il n’est pas possible d’envisager une solution qui irait du

point rouge au point bleu en changeant d'état discret d'abord u puis v (ou le contraire) : la transition est donc nécessairement synchrone. Abandonner une approche *asynchrone* aurait de grosses conséquences sur les propriétés du cadre de modélisation et nous avons fait le choix d'ignorer ce cas particulier au profit de la simplicité. La suite de cette section démontre que ce choix n'est pas déraisonnable, car des situations du type de la figure 5.11 droite imposent des graphes d'interaction pour le moins improbables biologiquement.

Rappelons qu'un mur noir impose une auto-inhibition sur la variable concernée au seuil correspondant au mur. Par conséquent, *toutes* les variables concernées par l'hypercoin doivent avoir une auto-inhibition sur le seuil concerné, et cette auto-inhibition doit être effective lorsque *toutes les autres* variables sont à vitesse nulle (stables). En particulier, pour obtenir l'un des graphes d'états de la figure 5.11, il est nécessaire d'avoir les deux auto-inhibitions de la figure 5.12 haut. De plus, pour obtenir la « fuite » de l'état $(n-1, m-1)$ de la figure 5.11 droite, il est nécessaire que u inhibe v au même seuil que sa propre auto-inhibition et que, dans le même temps v inhibe u également au même seuil que sa propre inhibition, ce qui conduit à la figure 5.12 bas.

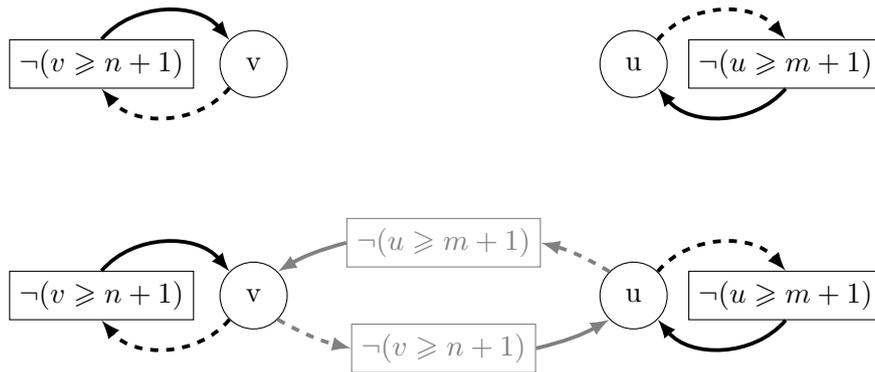


FIGURE 5.12 – Réseau à 2 variables dont la dynamique est un hypercoin. Les multiplexes gris peuvent ne pas être présents.

Biologiquement, il est très peu probable que deux seuils soient égaux, même s'ils peuvent, le cas échéant, différer de très peu en valeurs réelles. Même dans le cas des opérons, par exemple, pour lesquels la transcription des gènes est simultanée, nous savons que la plupart du temps, l'épissage (prétraitement de l'ARN messager) et la traduction séparent les seuils. En pratique, dans la quasi-totalité des graphes d'interactions génétiques, tous les seuils sortant d'une même variable sont différents. *A fortiori*, il est fort peu probable que plusieurs variables s'entre-inhibent précisément au même seuil que leur propre auto-inhibition, et que toutes ces inhibitions soient fonctionnelles lorsque toutes les autres variables sont stables.

Au vu de la faible probabilité d'apparition de tels réseaux et des problèmes que générerait la gestion de tels cas particuliers, il semblait plus raisonnable de préserver la simplicité de la définition 19.

5.5 Lien avec le cadre discret de René Thomas

Le formalisme hybride étant une extension du formalisme discret introduit par Thomas, nous avons dans un premier temps posé une variante de la définition 13 qui était totalement compatible avec cette théorie : pour toutes les variables v et pour toutes les ressources ω , il devait nécessairement y avoir un état n_0 dans lequel la célérité C_{v,ω,n_0} était nulle. Comme on l'a déjà remarqué, cet état n_0 n'est autre que $K_{v,\omega}$. Dès lors, il existait une fonction d'abstraction qui permettait de convertir un réseau hybride en réseau de Thomas. Mieux même : puisque les célérités des valeurs n inférieures à celles du paramètre $K_{v,\omega} = n_0$ sont strictement positives ($C_{v,\omega,n} > 0$), celles supérieures sont strictement négatives ($C_{v,\omega,n} < 0$) et celles telles que $n = n_0$ qui ont la même valeur sont nulles ($C_{v,\omega,n_0} = 0$). Ainsi la connaissance du réseau discret de Thomas détermine le signe des célérités.

Par extension, au niveau des graphes d'états avec cette première version de la définition 13, il suffisait presque de ne retenir que la partie discrète de chaque état hybride et l'abstraction préservait la dynamique discrète modulo les conventions classiques introduites par Houssine Snoussi [64] :

- en effet, dans le cadre des équations différentielles linéaires par morceaux étudié par Houssine Snoussi, les transitions sont complètement déterministes et il n'existe plusieurs transitions sortantes d'un même état discret que parce que des trajectoires *différentes* franchissent les faces correspondantes. Il en est de même pour notre approche hybride. Si $C_{v,\omega,n} > 0$ (resp. $C_{v,\omega,n} < 0$) alors, l'abstraction de la dynamique produit une transition pour la variable v de n vers $n + 1$ (resp. n à $n - 1$) cf. figure 5.13.

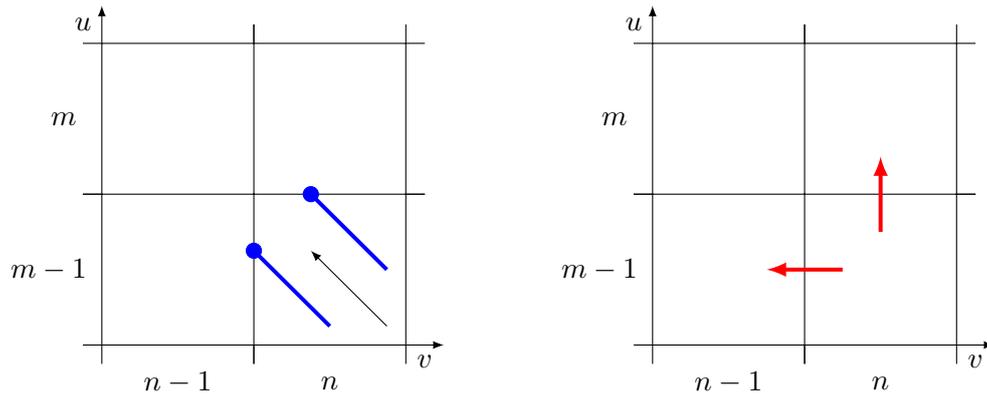


FIGURE 5.13 – À gauche : la célérité en v est négative et celle en u est positive. À droite : l'abstraction discrète produit par conséquent une transition vers l'état discret de gauche et une autre vers l'état discret du dessus.

- Dans le cas des murs noirs à un seuil n , la fonction d'abstraction construit à la fois une transition qui transforme v de n à $n - 1$ et une autre qui transforme v de $n - 1$ à n , alors qu'il n'existe pas de transition hybride, ni de trajectoire dans

les équations différentielles linéaires par morceaux, qui franchit le seuil entre n et $n - 1$. cf. figure 5.14.

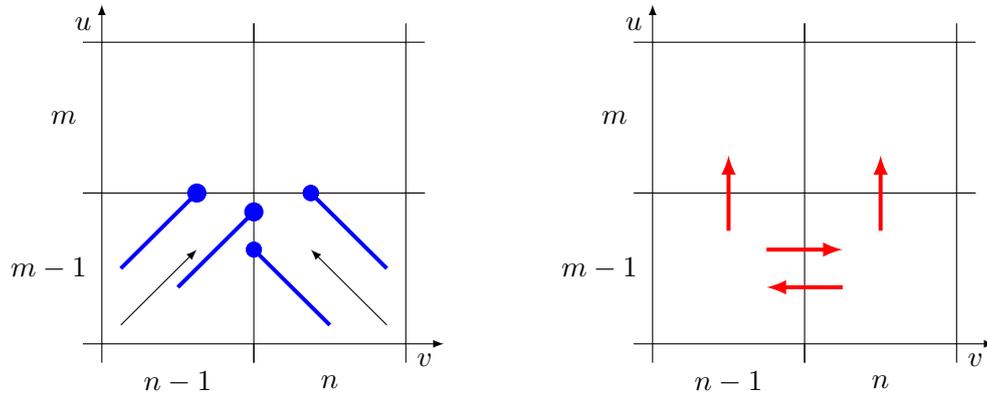


FIGURE 5.14 – Un mur noir génère une oscillation possible (artéfactuelle) entre deux états discrets.

- L'abstraction vers le discret augmente sensiblement le non-déterminisme des trajectoires. Cela a pour conséquence que certaines trajectoires discrètes ne possèdent aucune trajectoire hybride (ou par équations différentielles linéaires par morceaux) correspondante. Par exemple, dans la figure 5.15 droite, la trajectoire verticale qui conduit $n - 1$ à $n + 1$ en deux étapes ne correspond à aucune trajectoire hybride possible de la figure 5.15 gauche.

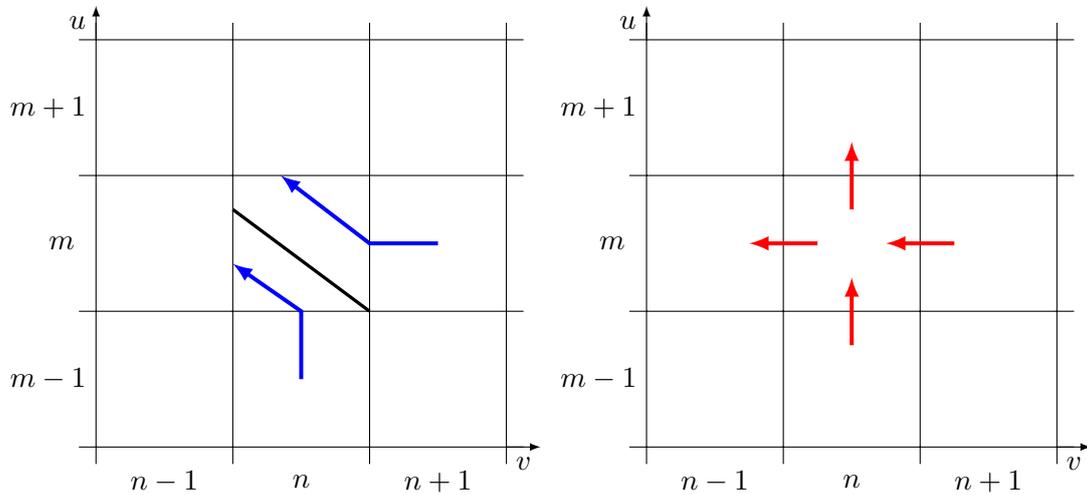


FIGURE 5.15 – Le non déterminisme induit par l'abstraction discrète augmente le nombre de chemins possibles.

Dans le cadre hybride, les états discrets stables sont de deux types :

- Les célérités de toutes les variables sont nulles à l'intérieur de l'état. Dans ce cas, de manière évidente, cet état correspond à un état régulier stable dans le formalisme de Thomas.
- Les célérités ne sont pas toutes nulles, mais la trajectoire ne pourra pas quitter

l'état courant. Par exemple, c'est le cas d'un état discret où toutes les trajectoires conduisent à un hypercoin.

La définition 13, dans sa version actuelle, permet que toutes les célérités soient strictement positives ou strictement négatives à v et ω fixés. Dans le cas positif, cela reviendrait intuitivement à poser $n_0 = b_v + 1$, et dans le cas négatif, à poser $n_0 = -1$. Il en résulte que ces cas de figure ne sont pas capturés par l'abstraction discrète définie ci-dessus, puisque $K_{v,\omega} = n_0$ sortirait de l'intervalle $[0 \dots b_v]$.

Ce qui a motivé cette extension est la surabondance d'exemples dans lesquels n_0 était égal à 0 ou à b_v , et pour lesquels l'ensemble des états $v = n_0$ n'avait aucun autre intérêt que de faire « glisser » les trajectoires au sein de ces « bords » de l'espace des états. Pire, cela rendait sans intérêt les variables booléennes, puisque dans l'un des deux états la vitesse doit être nulle.

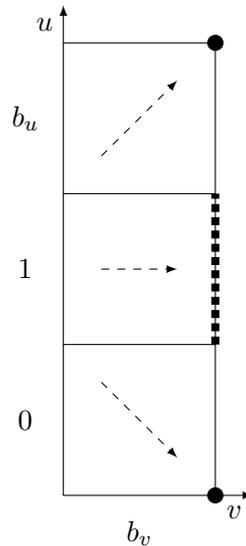


FIGURE 5.16 – Exemples de comportements d'une trajectoire dans le formalisme hybride lorsqu'une variable v a une célérité strictement positive au niveau de sa borne supérieure.

Selon la définition 13, pour toute variable v , si elle est dans un état η tel que $\eta(v) = 0$ ou $\eta(v) = b_v$ et que la célérité de v fait aller la trajectoire vers le bord de l'espace des états, celle-ci ne pourra pas sortir de cet état discret. Celle-ci devra glisser sur le bord. La figure 5.16 en présente des exemples. Dans les états $(b_v, 0)$ et (b_v, b_u) , les deux célérités ne sont pas nulles, mais quand la trajectoire arrivera dans le coin, elle ne pourra plus évoluer. Selon les principes que nous avons adoptés dans la première version, d'après le signe des célérités, il devrait y avoir des transitions entre cet état discret et l'extérieur de l'espace des états. Comme cela n'est pas possible, l'état est donc stable. Dans l'état $(b_v, 1)$, la célérité de u est nulle, alors que celle de v fait que la trajectoire atteint le bord de l'espace des états. Arrivée sur le bord, la trajectoire ne peut plus évoluer, donc, de la même façon que précédemment, cet état est stable.

5.6 Conclusion

Dans ce chapitre, nous avons présenté un nouveau cadre de modélisation hybride. Il est basé sur le formalisme de René Thomas [69], dont l'approche purement discrète permet, entre autres, d'utiliser des logiques temporelles ou des méthodes à base de contraintes pour une identification automatique des paramètres des modèles en fonction des propriétés biologiques connues. Dans le formalisme de Thomas, la notion de temps est uniquement chronologique, autrement dit il ne prend en compte que l'ordre des événements, mais pas les délais entre eux. Notre objectif était donc de définir une extension au cadre de modélisation de Thomas en y ajoutant la notion de délai.

Comme, nous l'avons vu, d'autres formalismes avaient déjà proposé des approches hybrides, mais celles-ci avaient le défaut principal de ne pas pouvoir prendre en compte les phénomènes d'accumulation. Seule la théorie développée conjointement par les équipes de Nice et de Nantes permettait de refléter ces phénomènes d'accumulation, mais au prix de transformations compliquées à chaque changement d'état discret, ce qui nuisait fortement à l'identification automatique des paramètres. En réaction à cela, l'approche développée dans ce chapitre a repris les idées de ces deux équipes, mais les a intégrées dans un formalisme plus simple grâce à la notion de célérité. Les célérités adoptent une philosophie proche de celle des paramètres discrets multivalués introduite par Houssine Snoussi, et les remplacent.

Cet apport permet aux paramètres, les célérités, de contenir à la fois la vitesse d'une variable dans un état et sa direction. Cette direction est étroitement liée, par une abstraction, à la valeur des paramètres de Thomas. En fait, les signes des célérités permettent de déduire les valeurs des paramètres de Thomas.

Sur le fond, une célérité est l'inverse d'un délai. Cette remarque est très utile, car les délais biologiques constituent une information aisément identifiable par expérimentation. Ceci rend l'identification des célérités plus facile. Par ailleurs, comme nous le verrons dans le chapitre 8, les définitions posées dans ce chapitre ont déjà été reprises par d'autres chercheurs des équipes de Nice et de Nantes qui ont étendu plusieurs techniques classiques d'identification des paramètres à notre cas hybride.

Pour aborder des difficultés malheureusement non résolues, nous nous devons de mentionner que le lien direct entre le formalisme des équations différentielles par morceaux et notre formalisme hybride n'a pas été établi. En effet, nous avons vu que, modulo certaines hypothèses non restrictives, il existe une abstraction de notre formalisme hybride vers le formalisme discret de Thomas. Il existe de même une abstraction du formalisme des équations différentielles linéaires par morceaux vers le formalisme de Thomas [64, 30]. Il nous avait semblé naturel, dans un premier temps, qu'il existe une abstraction du formalisme des équations linéaires par morceaux vers notre formalisme hybride, mais il semble que ce ne soit malheureusement pas le cas. L'idée de base que nous avons adoptée consis-

tait à « tirer une corde » à partir d'une trajectoire différentiable qui délimite la frontière entre plusieurs faces de sorties possibles d'un état discret (figure 5.17). Malheureusement, cette idée qui avait l'avantage de préserver les successions d'états discrets dans les trajectoires ne se généralise pas aux dimensions supérieures.

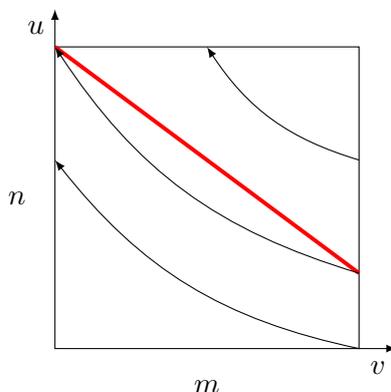


FIGURE 5.17 – Le formalisme hybride ne permet pas d'abstraire, du moins d'une manière simple, les formalismes par équations différentielles, essentiellement parce qu'on ne trouve pas d'approximation des célérités compatibles avec toutes les faces de sortie des états discrets.

Il n'en reste pas moins que le formalisme hybride fournit des modèles très prédictifs de réseaux génétiques avec délais, comme nous le démontrerons dans le chapitre 7, où nous présenterons l'application du formalisme défini dans ce chapitre sur le cycle circadien. Les résultats obtenus sont remarquablement satisfaisants par rapport à la simplicité du réseau hybride développé.

Chapitre 6

Modèle hybride circadien et ses propriétés remarquables

Le cycle circadien joue un rôle très important chez tous les mammifères. Par exemple, c'est un élément clé dans toutes les problématiques de chronopharmacologie. En particulier, il est connu pour être étroitement lié au cycle cellulaire et donc il a une place importante dans les problématiques liées au cancer.

Dans ce chapitre, nous donnerons d'abord une description des mécanismes responsables du cycle circadien. Puis, après avoir construit un modèle très simplifié de ce mécanisme, nous démontrerons la propriété qui énonce que dans un réseau à 2 variables contenant un comportement cyclique, la présence d'un glissement sur un des murs extérieurs est nécessaire et suffisante à l'obtention d'un attracteur cyclique. Enfin nous déterminerons les contraintes qui permettent d'assurer l'une des caractéristiques principales du cycle circadien : une oscillation entretenue même en absence de lumière et avec une période fixée.

6.1 Le cycle circadien

6.1.1 Introduction

Le cycle circadien est l'adaptation du vivant à l'alternance jour/nuit (du grec *circa*, autour et *dies*, jour) afin d'optimiser la physiologie de l'organisme en permettant au métabolisme d'anticiper les variations quotidiennes de l'environnement. Parmi les adaptations classiques, l'horloge contrôle la température, la réparation des tissus, le sommeil, *etc.* Ce cycle compte parmi les mécanismes physiologiques les plus importants (43% des gènes sont régulés par ce cycle chez la souris [79]). Au-delà des mammifères, l'immense majorité des êtres vivants étant soumis à l'oscillation jour/nuit, le cycle circadien est omniprésent dans le vivant.

L'horloge circadienne la plus primitive se trouve chez les cyanobactéries (bactéries photosynthétiques). Cette horloge ne contient aucun gène, seulement 3 protéines dont une qui est phosphorylée et déphosphorylée plusieurs fois. Pour faire un cycle complet, ce système met environ 24 heures, donc remplit une fonction circadienne.

Chez l'Homme, l'horloge circadienne va réguler certains processus tels que le sommeil ou les processus liés aux fonctions rénales [18]. Les causes de pathologies dues à une horloge circadienne non fonctionnelle sont nombreuses [73, 66]

Dans tous les tissus, chaque cellule possède sa propre horloge circadienne qui, en l'absence de stimuli extérieurs, oscille avec sa propre période (entre 20h et 26h). Au sein d'un tissu, la communication cellulaire permet une synchronisation de toutes les cellules du tissu *via* un signal de régulation net [78]. Cependant, la synchronisation d'un tissu ne signifie pas forcément une synchronisation vis-à-vis de l'environnement. Cette synchronisation est assurée par les *zeitgebers*, « donneurs de temps » en allemand. Ces stimuli extérieurs

permettent à tout l'organisme de s'adapter à l'environnement, donc d'adapter sa période à exactement 24 heures. Chez les mammifères, le *zeitgeber* externe à l'organisme le plus important est la lumière, car c'est l'oscillation circadienne par excellence, mais ce n'est pas le seul. L'apport de nourriture, la température ou encore les rythmes sociaux ont aussi une influence sur le cycle.

Chez les mammifères, bien que chaque cellule possède son réseau génétique de l'horloge circadienne, il existe deux types d'horloges :

- Une horloge centrale qui est assurée par le noyau suprachiasmatique, un groupe de neurones situé dans l'hypothalamus. C'est elle qui reçoit l'information lumineuse *via* la voie rétino-hypothalamique, qui transmet l'information lumineuse convertie par l'œil au noyau suprachiasmatique dans lequel une voie de signalisation va être activée pour synchroniser l'horloge.

L'horloge centrale a également le rôle de « chef d'orchestre » de tout l'organisme. *Via* la synthèse d'hormones, elle envoie des signaux de synchronisation (en particulier du cortisol [7]) dans tout l'organisme pour coordonner l'ensemble des tissus.

- Des horloges périphériques qui vont réguler la physiologie du tissu dans lequel elles se situent. En l'absence de signaux provenant du noyau suprachiasmatique, elles oscillent avec leur propre période et leur propre phase.

C'est un mécanisme biologique très étudié pour son rôle important dans bon nombre de pathologies, dans la chronopharmacologie, mais aussi pour d'autres domaines tels que le travail à horaires décalés. Jusqu'à maintenant, il a été essentiellement modélisé avec des systèmes d'équations différentielles [8, 33, 48, 49, 37] et avec un nombre de variables assez conséquent.

6.1.2 L'horloge moléculaire

À partir de maintenant, nous allons nous intéresser au cycle circadien des mammifères, et plus particulièrement aux mécanismes moléculaires responsables de ce cycle. Le cœur de cette horloge, également appelé « *pacemaker* » est un réseau génétique. Il est composé d'une boucle négative principale qui permet l'oscillation et de boucles périphériques annexes régulant plus finement la période de l'horloge. La figure 6.1 contient un schéma de *pacemaker*.

La boucle principale est composée de 7 gènes : les gènes *Per* (*Per1*, *Per2*, *Per3*), les gènes *Cry* (*Cry1*, *Cry2*), *Bmal1* et *Clock* (ou *Npas2*). Les protéines BMAL1 et CLOCK forment un complexe qui active la transcription des gènes *Per* et *Cry* (BMAL1-CLOCK est le facteur de transcription de *Per* et de *Cry*). Une fois synthétisées dans le cytoplasme, PER et CRY forment des combinaisons d'hétérodimères PER-CRY et s'accumulent. Ils sont ensuite phosphorylés par la caséine kinase CKI ϵ , ce qui permet aux complexes PER-CRY de migrer vers le noyau [76]. Dans le noyau, ils forment un complexe avec BMAL1-

CLOCK pour inhiber la transcription de leurs propres gènes *Per* et *Cry*.

L'effet de la lumière (*via* la voie rétino-hypothalamique) est l'activation de la transcription des gènes *Per* par l'acétylation de ces gènes [12].

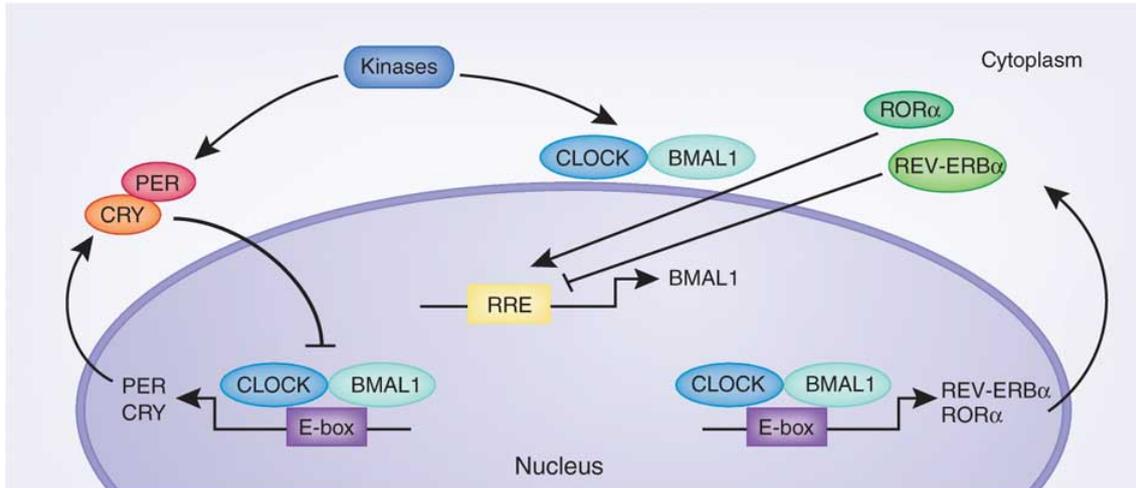


FIGURE 6.1 – Représentation des interactions entre les composants du « pacemaker » de l'horloge moléculaire.

Il y a donc 2 groupes d'acteurs :

- les activateurs constitués des facteurs de transcription BMAL1-CLOCK et des gènes *Per* et *Cry*,
- les inhibiteurs constitués des complexes PER-CRY lorsqu'ils ont été transportés dans le noyau.

Ces deux groupes d'acteurs constituent une boucle négative. Nous avons donc les éléments minimaux pour avoir une oscillation.

Les boucles périphériques annexes sont composées, d'une part, d'une activation de *Reverba* par le complexe BMAL1-CLOCK suivie d'une inhibition des gènes *Bmal1*, *Per1-3* et *Cry1-2* par REVERB α , et d'autre part, d'une activation de *Rora* par le complexe BMAL1-CLOCK suivie d'une inhibition des mêmes gènes par ROR α [38]. Ces deux boucles ont donc les mêmes cibles sur l'horloge principale de telle sorte que, dans notre contexte, le complexe BMAL1-CLOCK n'est jamais sous-représenté et c'est l'action du complexe PER-CRY sur BMAL1-CLOCK qui inhibe le premier groupe d'acteurs. Au sein du noyau suprachiasmatique et pour l'étude du rôle de la lumière qui nous intéresse, il semble que ces deux boucles n'aient qu'une fonction de renforcement de la boucle principale. Notre objectif étant de constituer un modèle simple de l'horloge, nous nous concentrerons uniquement sur le cœur de l'horloge constitué des deux groupes d'acteurs susmentionnés.

6.2 Modèles et propriété remarquable

6.2.1 Le schéma des modèles

Comme déjà mentionné, nous pouvons concevoir un modèle à deux variables qui reflète assez fidèlement les interactions entre les deux groupes d'acteurs. Il est inspiré de [21] qui présentait à la fois une version avec équations différentielles à quatre variables et une version discrétisée à deux variables.

Notre modèle contient les deux variables g et pc :

- g représente le premier groupe d'acteurs, c'est-à-dire principalement les gènes *Per1,2,3* et *Cry1,2* qui ont une action positive sur pc
- pc représente le second groupe d'acteurs, c'est-à-dire les complexe PER-CRY présents dans le noyau. pc va donc être inhibiteur de g pour représenter l'inhibition du complexe PER-CRY sur le facteur de transcription BMAL1-CLOCK.

Nous ajoutons également un zeitgeber L qui représente la lumière. Pour représenter l'action positive de l'acétylation des gènes *Per*, cette variable active g . Nous avons donc le réseau de la figure 6.2.

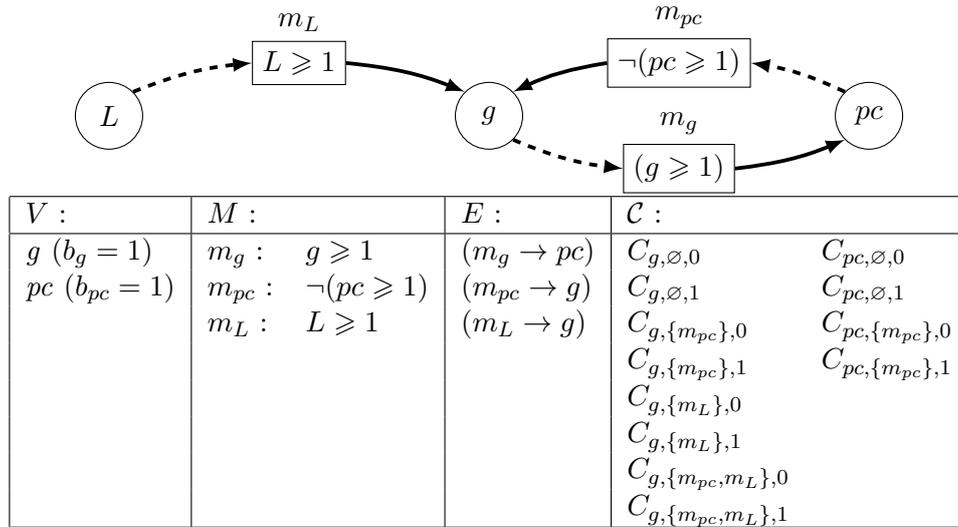


FIGURE 6.2 – Modèle simple de l'horloge circadienne. Haut : graphe d'interaction du modèle. Bas : le quadruplet (V, M, E, \mathcal{C}) de ce réseau.

Il faut maintenant identifier un jeu de célérités permettant d'obtenir des comportements compatibles avec celui de l'horloge des mammifères. Une des caractéristiques principales de l'horloge circadienne est la capacité à osciller autour de 24 heures sans aucune influence de l'environnement. Nous savons également que cette oscillation autonome est robuste à des perturbations importantes. Commençons par identifier des paramètres qui permettent ce comportement. Les contraintes obtenues vont donc être indépendantes du zeitgeber L . Ce sera l'objet des deux prochaines sections, mais commençons par présenter une propriété spécifique à un modèle à 2 dimensions.

6.2.2 Propriété remarquable en dimension 2

En accord avec le graphe de la figure 6.2, les célérités qui s'appliquent pour les variables g et pc dans chacun des quatre états possibles se répartissent comme indiqué dans la figure 6.3.

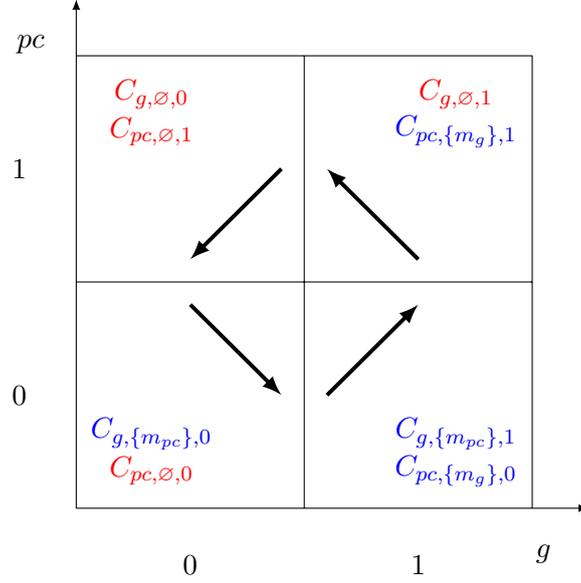


FIGURE 6.3 – Répartition des célérités et leur signe (les rouges sont négatives et les bleues positives).

Nous utiliserons également les notations ci-dessous dans la suite de cette sous-section.

Notation 2. On note respectivement $slope_{00}, slope_{01}, slope_{10}, slope_{11} \in \mathbb{R}_+^*$ les pentes des trajectoires à l'intérieur des états discrets $(0, 0), (0, 1), (1, 0), (1, 1)$ respectivement.

$$\begin{aligned}
 - slope_{00} &= \frac{-C_{pc,\emptyset,0}}{C_{g,\{m_{pc}\},0}} > 0 & - slope_{10} &= \frac{C_{pc,\{m_g\},0}}{C_{g,\{m_{pc}\},1}} > 0 \\
 - slope_{01} &= \frac{C_{pc,\emptyset,1}}{C_{g,\emptyset,0}} > 0 & - slope_{11} &= \frac{C_{pc,\{m_g\},1}}{-C_{g,\emptyset,1}} > 0
 \end{aligned}$$

Le fait que nos célérités soient constantes dans un état donné permet d'énoncer la proposition suivante.

Proposition 1. *Étant donné un réseau hybride $R = (V, M, E, C)$ où $|V| = 2$ et les bornes des deux variables sont égales à 1, il existe un unique cycle limite attracteur si et seulement si il existe une trajectoire cyclique qui glisse sur au moins un mur extérieur.*

Démonstration. Condition nécessaire : Si la trajectoire cyclique ne glisse sur aucun des murs, comme toutes les pentes sont constantes dans chacun des états discrets, il y a une infinité de trajectoires cycliques parallèles différentes (Figure 6.4-A). Il y a donc nécessairement au moins un glissement pour avoir un unique cycle limite.

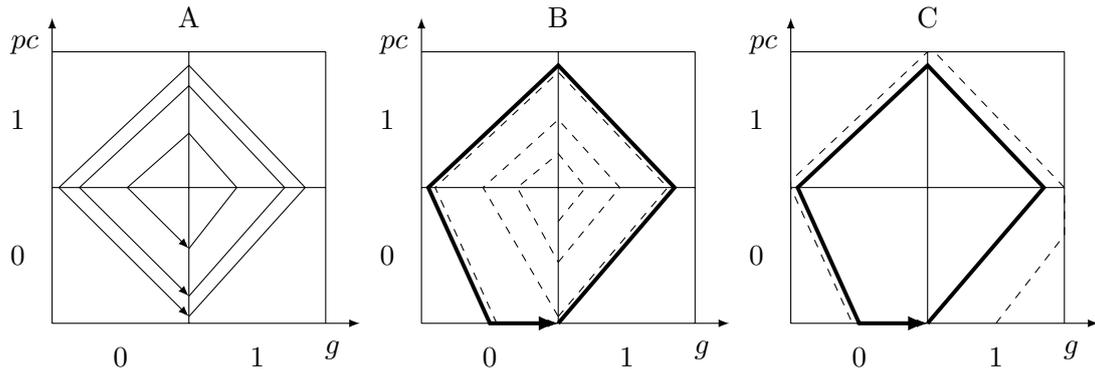


FIGURE 6.4 – graphe d'états (A) avec une infinité de trajectoires cycliques. (B) avec une trajectoire à l'intérieur du cycle limite. (C) avec une trajectoire extérieure au cycle

Condition suffisante : s'il existe une trajectoire qui glisse sur un mur extérieur, elle est unique et attractive, car :

- Pour tout état hybride à l'intérieur du cycle (Figure 6.4-B), sa trajectoire après une rotation (en pointillé) mène à un état strictement plus proche du cycle (la trajectoire est divergente). L'écart vers l'extérieur est proportionnel au rayon de rotation en raison du théorème de Thalès. Cet écart est donc croissant. Il en résulte que, après un certain nombre de rotations, la trajectoire finit par atteindre la trajectoire cyclique (trait plein épais).
- De la même manière, pour tout état à l'extérieur du cycle (Figure 6.4-C), la trajectoire reste à l'extérieur, car, à vitesse constante, les trajectoires ne peuvent pas se croiser (trajectoires parallèles). Par conséquent, la trajectoire va glisser sur les murs extérieurs et atteindre le cycle limite en, au plus, une rotation.

□

Pour notre modèle de l'horloge circadienne, nous voulons une oscillation soutenue d'environ 24 heures qui soit robuste à des perturbations. Comme nous sommes en 2 dimensions, nous savons donc qu'il faut au moins un glissement sur un des murs extérieurs.

6.3 Détermination des « pentes » des modèles

Afin de calculer les contraintes pour l'obtention d'un cycle limite circadien, nous allons traiter chaque cas dans lequel un glissement est fixé. La figure 6.5 fournit un bilan des différents types de trajectoires qui peuvent être observées en fonction de la valeur des pentes dans chaque état discret (bleu : sans glissement, rouge : avec glissement). Elle est utilisée dans toute la suite de cette section.

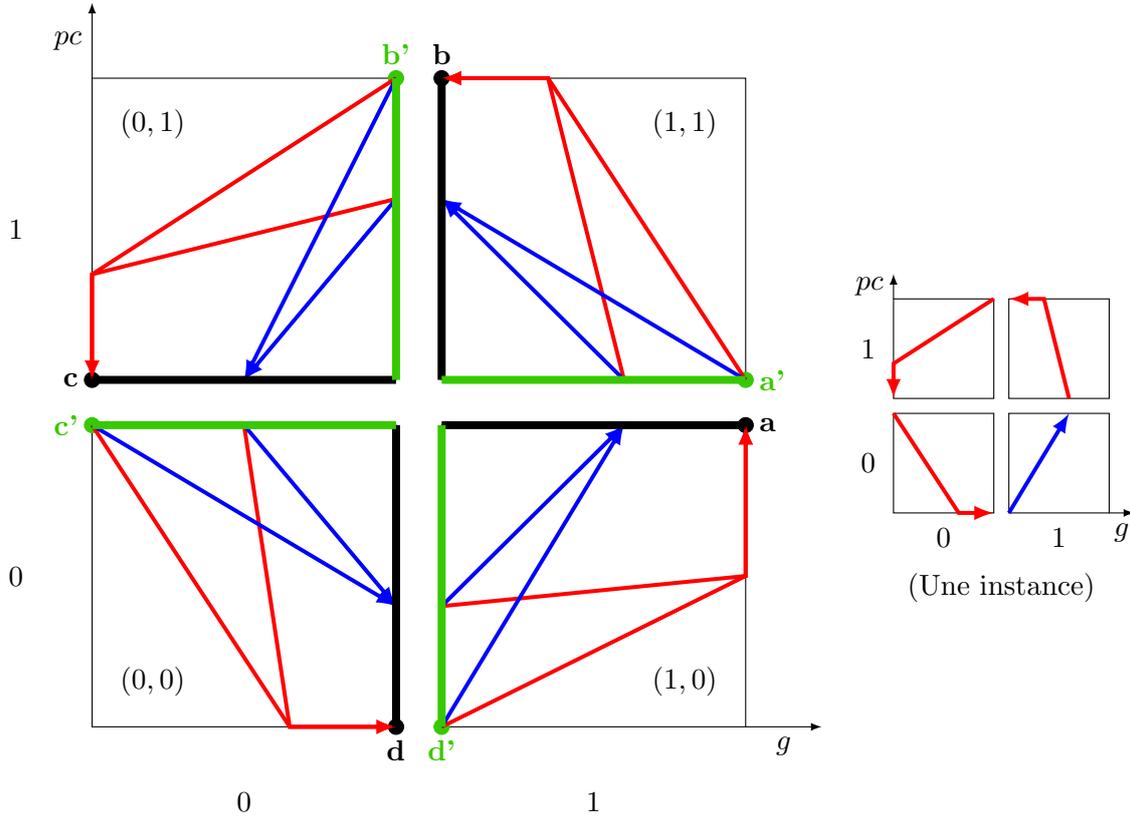


FIGURE 6.5 – Tous les types de trajectoires possibles dans un cycle limite en 2 dimensions. Les flèches bleues représentent les trajectoires qui ne touchent pas un des murs extérieurs, et les flèches rouges celles qui glissent sur ces murs. Les zones indiquées en noir gras, \mathbf{a} , \mathbf{b} , \mathbf{c} et \mathbf{d} représentent l'ensemble des états hybrides d'arrivées possibles depuis chacun des états discrets et les zones en vert gras \mathbf{a}' , \mathbf{b}' , \mathbf{c}' et \mathbf{d}' représentent leurs équivalents après le passage du seuil adjacent. Les points noirs gras indiquent les points de passage à l'issue d'un glissement. La petite figure de droite représente une instance de ces trajectoires où les glissements ont lieu au sein des états $(0,0)$, $(1,1)$ et $(0,1)$.

6.3.1 Glissement imposé en $(0,0)$

Le glissement étant imposé sur $(0,0)$, la trajectoire dans l'état $(1,0)$ part nécessairement du point vert de \mathbf{d}' , c'est-à-dire l'état tel que $\pi_{d'}(g) = \pi_{d'}(pc) = 0$. On cherche maintenant à calculer l'état sur lequel la trajectoire arrive après avoir traversé l'état $(1,0)$ (au sein de la zone \mathbf{a} sur la figure 6.5). On sait que $\eta_a = (1,0)$ et $\pi_a(pc) = 1$. Pour calculer la valeur de $\pi_a(g)$, il y a deux cas différents sont à considérer :

- Si la trajectoire touche la borne supérieure de g , elle ne pourra plus évoluer avant de changer d'état discret (glissement supplémentaire, trajectoire rouge). Nous avons donc $\pi_a(g) = 1$.
- Si pc est le premier à atteindre son seuil, g ne touche pas sa borne (trajectoire bleue). Nous avons donc $\pi_a(g) = \frac{1}{slope_{10}}$.

En conclusion $\pi_a(g) = \inf\left(\frac{1}{slope_{10}}, 1\right)$

Calculons maintenant l'état d'arrivée après la traversée de l'état $(1, 1)$ parmi les états de **b**. Nous savons que $\eta_b = (1, 1)$ et $\pi_b(g) = 0$. Il y a à nouveau plusieurs cas :

- Si la trajectoire touche la borne supérieure de pc (trajectoires rouges en $(1, 1)$), il y a glissement donc $\pi_b(pc) = 1$
- Si la trajectoire ne touche pas la borne supérieure de pc , mais qu'elle avait touché celle de g dans l'état $(1, 0)$ (trajectoire rouge en $(1, 0)$ et bleu en $(1, 1)$), alors on a $\pi_b(pc) = slope_{11}$.
- Si la trajectoire n'a touché aucune borne (bleu en $(1, 0)$ et en $(1, 1)$), on a $\pi_b(pc) = \frac{slope_{11}}{slope_{10}}$

En conclusion $\pi_b(pc) = \inf\left(\frac{slope_{11}}{slope_{10}}, slope_{11}, 1\right)$

Calculons maintenant l'état d'arrivée après la traversée de l'état $(0, 1)$ parmi les états de **c**. Nous savons que $\eta_c = (0, 1)$ et $\pi_c(pc) = 0$. Il y a plusieurs cas pour $\pi_c(g)$:

- Si la trajectoire touche la borne inférieure de g (trajectoires rouge en $(0, 1)$), il y a glissement donc $\pi_c(g) = 0$
- Si la trajectoire ne touche pas la borne inférieure de g , mais qu'elle a touché la borne supérieure de pc dans l'état $(1, 1)$ (trajectoires rouges en $(1, 1)$ et bleue en $(0, 1)$), on a $\pi_c(g) = \frac{1}{slope_{01}}$.
- Si la trajectoire ne touche pas la borne inférieure de g , mais qu'elle a touché la borne supérieure de g dans l'état $(1, 0)$ (trajectoire rouge en $(1, 0)$ et bleues en $(1, 1)$ et $(0, 1)$), on a $\pi_c(g) = \frac{slope_{11}}{slope_{01}}$
- Si la trajectoire n'a touché aucune borne (uniquement des trajectoires bleues), on a $\pi_c(g) = \frac{slope_{11}}{slope_{01} \times slope_{10}}$

En conclusion $\pi_c(g) = \inf\left(\frac{slope_{11}}{slope_{10} \times slope_{01}}, \frac{slope_{11}}{slope_{01}}, \frac{1}{slope_{01}}, 1\right)$

Afin de fermer la trajectoire, il faut que toutes les trajectoires possibles glissent contre la borne inférieure de pc . Autrement dit, il faut qu'il y ait une trajectoire rouge dans l'état $(0, 0)$:

- Si la trajectoire a glissé dans $(0, 1)$ (trajectoire rouge dans $(0, 1)$, puis rouge dans $(0, 0)$), on doit avoir $slope_{00} > 1$
- Si la trajectoire n'a pas glissé dans $(0, 1)$, mais qu'elle a glissé dans $(1, 1)$ (trajectoire rouge dans $(1, 1)$, bleue dans $(0, 1)$ puis rouge dans $(0, 0)$), on doit avoir $\frac{slope_{00}}{slope_{01}} > 1$ donc $slope_{00} > slope_{01}$
- Si la trajectoire n'a glissé ni dans $(0, 1)$, ni dans $(1, 1)$, mais qu'elle a glissé dans $(1, 0)$ (trajectoire rouge dans $(1, 0)$ et bleue dans $(1, 1)$ et $(0, 1)$, puis rouge dans $(0, 0)$), on a $\frac{slope_{00} \times slope_{11}}{slope_{01}} > 1$ donc $slope_{00} \times slope_{11} > slope_{01}$
- Si la trajectoire n'a glissé nulle part (trajectoire bleue dans $(1, 0)$, $(1, 1)$ et $(0, 1)$, puis rouge dans $(0, 0)$), on doit avoir $\frac{slope_{00} \times slope_{11}}{slope_{10} \times slope_{01}} > 1$ donc $slope_{00} \times slope_{11} > slope_{01} \times slope_{10}$

In fine, avec un glissement imposé en $(0, 0)$, on obtient donc la contrainte suivante :

$$\begin{aligned}
& slope_{00} > 1 && \text{(glissement sur } (0, 1)) \\
\vee & slope_{00} > slope_{01} && \text{(glissement sur } (1, 1)) \\
\vee & slope_{00} \times slope_{11} > slope_{01} && \text{(glissement sur } (1, 0)) \\
\vee & slope_{00} \times slope_{11} > slope_{01} \times slope_{10} && \text{(aucun glissement supplémentaire)}
\end{aligned}$$

Dans cette disjonction, le nombre d'atomes parmi les 3 premiers qui sont vrais indique le nombre de glissements supplémentaires du cycle limite.

6.3.2 Glissement imposé sur (1, 0)

Comme le glissement est imposé sur (1, 0), la trajectoire dans l'état (1, 1) part nécessairement du point vert de \mathbf{a}' , c'est-à-dire l'état dans tel que $\pi_{a'}(g) = 1$ et $\pi_{a'}(pc) = 0$. De la même façon que précédemment, mais avec une rotation d'un quart de tour, nous pouvons déterminer les contraintes sur les pentes :

$$\begin{aligned}
& \pi_b(pc) = \inf(slope_{11}, 1) \\
& \pi_c(g) = \inf\left(\frac{slope_{11}}{slope_{01}}, \frac{1}{slope_{01}}, 1\right) \\
& \pi_d(pc) = \inf\left(\frac{slope_{11} \times slope_{00}}{slope_{01}}, \frac{slope_{00}}{slope_{01}}, slope_{00}, 1\right) \\
& \pi_a(g) = \inf\left(\frac{slope_{00} \times slope_{11}}{slope_{01} \times slope_{10}}, \frac{slope_{00}}{slope_{01} \times slope_{10}}, \frac{slope_{00}}{slope_{10}}, \frac{1}{slope_{10}}\right) > 1
\end{aligned}$$

Les contraintes pour un cycle limite dans lequel un glissement est imposé sur (1, 0) sont donc les suivantes :

$$\begin{aligned}
& slope_{10} > 1 && \text{(glissement sur } (0, 0)) \\
\vee & slope_{00} > slope_{10} && \text{(glissement sur } (0, 1)) \\
\vee & slope_{00} > slope_{01} \times slope_{10} && \text{(glissement sur } (1, 1)) \\
\vee & slope_{00} \times slope_{11} > slope_{01} \times slope_{10} && \text{(aucun glissement supplémentaire)}
\end{aligned}$$

6.3.3 Glissement imposé sur (1, 1)

Comme le glissement est imposé sur (1, 1), la trajectoire dans l'état (0, 1) part nécessairement du point vert de \mathbf{b}' , c'est-à-dire l'état dans tel que $\pi_{b'}(g) = \pi_{b'}(pc) = 1$. De la même façon que précédemment, nous pouvons déterminer les contraintes sur les pentes :

$$\begin{aligned}
& \pi_c(g) = \inf\left(\frac{1}{slope_{01}}, 1\right) \\
& \pi_d(pc) = \inf\left(\frac{slope_{00}}{slope_{01}}, slope_{00}, 1\right) \\
& \pi_a(g) = \inf\left(\frac{slope_{00}}{slope_{10} \times slope_{01}}, \frac{slope_{00}}{slope_{10}}, \frac{1}{slope_{10}}, 1\right) \\
& \pi_b(pc) = \inf\left(\frac{slope_{00} \times slope_{11}}{slope_{01} \times slope_{10}}, \frac{slope_{00}}{slope_{01} \times slope_{10}}, \frac{slope_{00}}{slope_{10}}, \frac{1}{slope_{10}}\right) > 1
\end{aligned}$$

Les contraintes pour un cycle limite dans lequel un glissement est imposé sur (1,1) sont donc suivantes :

$$\begin{aligned}
& slope_{11} > 1 && \text{(glissement sur (1,0))} \\
\vee & slope_{11} > slope_{10} && \text{(glissement sur (0,0))} \\
\vee & slope_{00} \times slope_{11} > slope_{10} && \text{(glissement sur (0,1))} \\
\vee & slope_{00} \times slope_{11} > slope_{01} \times slope_{10} && \text{(aucun glissement supplémentaire)}
\end{aligned}$$

6.3.4 Glissement imposé sur (0,1)

Comme le glissement est imposé sur (0,1), la trajectoire dans l'état (0,0) part nécessairement du point vert de \mathbf{c}' , c'est-à-dire l'état dans tel que $\pi_{c'}(g) = 0$ et $\pi_{c'}(pc) = 1$. De la même façon que précédemment, nous pouvons déterminer les contraintes sur les pentes :

$$\begin{aligned}
- & \pi_d(pc) = inf(slope_{00}, 1) \\
- & \pi_a(g) = inf\left(\frac{slope_{00}}{slope_{10}}, \frac{1}{slope_{10}}, 1\right) \\
- & \pi_b(pc) = inf\left(\frac{slope_{11} \times slope_{00}}{slope_{10}}, \frac{slope_{11}}{slope_{10}}, slope_{11}, 1\right) \\
- & \pi_c(g) = inf\left(\frac{slope_{00} \times slope_{11}}{slope_{01} \times slope_{10}}, \frac{slope_{00}}{slope_{01} \times slope_{10}}, \frac{slope_{00}}{slope_{10}}, \frac{1}{slope_{10}}\right) > 1
\end{aligned}$$

Les contraintes pour un cycle limite dans lequel un glissement est imposé sur (0,1) sont donc les suivantes :

$$\begin{aligned}
& slope_{01} < 1 && \text{(glissement sur (1,1))} \\
\vee & slope_{11} > slope_{01} && \text{(glissement sur (1,0))} \\
\vee & slope_{11} > slope_{10} \times slope_{01} && \text{(glissement sur (0,0))} \\
\vee & slope_{00} \times slope_{11} > slope_{01} \times slope_{10} && \text{(aucun glissement supplémentaire)}
\end{aligned}$$

Au bilan

Dans toutes les conditions de cette section 6.3, la formule $slope_{00} \times slope_{11} > slope_{01} \times slope_{10}$ est présente. Cette formule est vraie s'il n'y a qu'un seul glissement dans la trajectoire, quel que soit le mur sur lequel il se situe : elle constitue donc la condition nécessaire et suffisante pour avoir un cycle limite (cf. section 6.2.2). Chacune des autres contraintes représente un cas possible de trajectoire (combinaisons d'états dans lesquels il y a glissement imposé sur une des bornes).

Chacun des atomes obtenus précédemment représente une contrainte sur les pentes entre deux glissements successifs. Par exemple, $(slope_{01} > 1)$ est la contrainte correspon-

nant aux trajectoires avec glissement au sein de $(0, 1)$ en sortant d'un glissement en $(1, 1)$. Afin d'avoir les contraintes sur toutes les pentes, il faut combiner les contraintes : si on veut des glissements sur $(0, 0)$, $(0, 1)$ et $(1, 1)$ comme dans la figure 6.6, les contraintes seront :

- $(0, 1) \rightarrow (0, 0) : slope_{00} > 1$
- $(1, 1) \rightarrow (0, 1) : slope_{01} > 1$
- $(0, 0) \rightarrow (1, 0) \rightarrow (1, 1) : slope_{11} > slope_{10}$

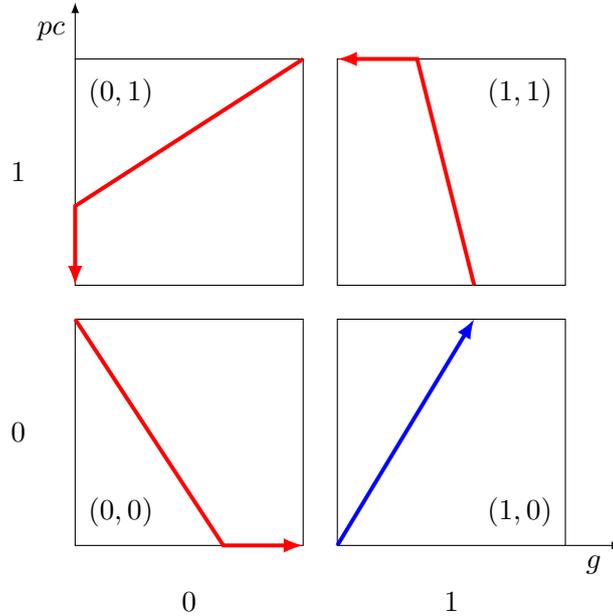


FIGURE 6.6 – Exemple de trajectoire avec glissements en $(0, 0)$, $(1, 1)$ et $(0, 1)$.

Nous connaissons maintenant les contraintes sur chacune des pentes afin d'obtenir un cycle limite avec n'importe quelle combinaison de glissements. Il faut également identifier des contraintes sur les célérités qui permettront d'obtenir une période donnée précise. Pour cela, il faut que la somme des temps passés dans chacun des états discrets soit égale à la période que l'on souhaite.

6.4 Détermination des célérités des modèles

Maintenant que nous connaissons les contraintes sur les pentes, cela nous donne une contrainte de proportionnalité entre les deux célérités qui s'appliquent dans chacun des états. On rappelle que la célérité peut être obtenue à partir du délai pour passer d'un bord à l'autre d'un état discret (c'est l'inverse du délai). La connaissance des pentes nous permet donc de déterminer les deux célérités de chaque état avec un seul délai par état. Il suffit par conséquent de savoir combien de temps est passé dans chaque état au cours du cycle limite.

Introduisons donc les délais δ_{00}^g , δ_{10}^{pc} , δ_{11}^g et δ_{01}^{pc} qui sont respectivement les délais de

passage dans les états $(0,0)$, $(1,0)$, $(1,1)$ et $(0,1)$. Comme on va le voir plus loin et conformément à la figure 6.5, il est plus judicieux de considérer, par exemple, le délai δ_{00}^g que le délai δ_{00}^{pc} . En effet, en $(0,0)$, pc peut « buter » sur un mur extérieur (trajectoires rouges), alors que g va nécessairement changer de valeur sans mur noir. Le délai δ_{00}^g facilitera donc les calculs, car il ne sera pas nécessaire d'introduire une expression algébrique pour un glissement en g . En revanche, le point d'entrée dans l'état $(0,0)$ devra être pris en compte.

De la même façon que pour les pentes, nous calculons, dans les sous-sections suivantes, les contraintes pour tous les cas de glissement.

6.4.1 Glissement imposé sur $(0,0)$

En accord avec la figure 6.5, un glissement sur $(0,0)$ fournit le point de départ dans l'état $(1,0)$. Le plus simple est donc de partir de cet état. Dans cet état, la trajectoire part forcément du coin inférieur gauche ($\pi_a(g) = 0$ et $\pi_a(pc) = 0$). C'est nécessairement pc qui change d'état discret en premier, quelle que soit la trajectoire. Nous avons ainsi $\delta_{10}^{pc} = \frac{1}{C_{pc,\rho((1,0),pc),0}}$.

Ensuite, dans l'état $(1,1)$:

- si la trajectoire de l'état $(1,0)$ a touché le mur extérieur (trajectoire rouge dans la figure 6.5), alors $\delta_{11}^g = \frac{1}{C_{g,\rho((1,1),g),1}}$.
- si elle n'a pas touché (trajectoire bleue), alors $\delta_{11}^g = \frac{1}{C_{g,\rho((1,1),g),1}} \times \frac{1}{slope_{10}}$

Donc finalement $\delta_{11}^g = \frac{1}{C_{g,\rho((1,1),g),1}} \times \inf\left(\frac{1}{slope_{10}}, 1\right)$.

Dans le même esprit, pour l'état $(0,1)$,

- si la trajectoire a touché un mur extérieur (trajectoire rouge) dans l'état $(1,1)$, alors $\delta_{01}^{pc} = \frac{1}{C_{pc,\rho((0,1),pc),1}}$.
- Si elle n'a pas touché (trajectoire bleue) en $(1,1)$ mais qu'elle l'a touché en $(1,0)$, alors $\delta_{01}^{pc} = \frac{1}{C_{pc,\rho((0,1),pc),1}} \times slope_{11}$.
- Enfin, si elle n'a pas touché de mur, alors $\delta_{01}^{pc} = \frac{1}{C_{pc,\rho((0,1),pc),1}} \times \frac{slope_{11}}{slope_{10}}$.

Donc finalement, $\delta_{01}^{pc} = \frac{1}{C_{pc,\rho((0,1),pc),1}} \times \inf\left(\frac{slope_{11}}{slope_{10}}, slope_{11}, 1\right)$.

Pour terminer la boucle avec l'état $(0,0)$,

- si le dernier mur extérieur touché est $(0,1)$ (trajectoire rouge), alors $\delta_{00}^g = \frac{1}{C_{g,\rho((0,0),g),0}}$.
- Si c'est $(1,1)$, alors $\delta_{00}^g = \frac{1}{C_{g,\rho((0,0),g),0}} \times \frac{1}{slope_{01}}$.
- Si le dernier mur atteint est dans $(1,0)$, alors $\delta_{00}^g = \frac{1}{C_{g,\rho((0,0),g),0}} \times \frac{slope_{11}}{slope_{01}}$
- s'il n'y a aucun glissement, alors $\delta_{00}^g = \frac{1}{C_{g,\rho((0,0),g),0}} \times \frac{slope_{11}}{slope_{10} \times slope_{01}}$.

Finalement, $\delta_{00}^g = \frac{1}{C_{g,\rho((0,0),g),0}} \times \inf\left(\frac{slope_{11}}{slope_{10} \times slope_{01}}, \frac{slope_{11}}{slope_{01}}, \frac{1}{slope_{01}}, 1\right)$.

Au bilan,

$$\begin{aligned}
- \delta_{10}^{pc} &= \frac{1}{C_{pc,\rho((1,0),pc),0}} \\
- \delta_{11}^g &= \frac{1}{C_{g,\rho((1,1),g),1}} \times \inf\left(\frac{1}{slope_{10}}, 1\right) \\
- \delta_{01}^{pc} &= \frac{1}{C_{pc,\rho((0,1),pc),1}} \times \inf\left(\frac{slope_{11}}{slope_{10}}, slope_{11}, 1\right) \\
- \delta_{00}^g &= \frac{1}{C_{g,\rho((0,0),g),0}} \times \inf\left(\frac{slope_{11}}{slope_{10} \times slope_{01}}, \frac{slope_{11}}{slope_{01}}, \frac{1}{slope_{01}}, 1\right)
\end{aligned}$$

Comme la somme $\delta_{10}^{pc} + \delta_{11}^g + \delta_{01}^{pc} + \delta_{00}^g$ doit être égale à la période souhaitée, nous obtenons :

$$\begin{aligned}
& \frac{1}{C_{pc,\rho((1,0),pc),0}} \\
& + \frac{1}{C_{g,\rho((1,1),g),1}} \times \inf\left(\frac{1}{slope_{10}}, 1\right) \\
& + \frac{1}{C_{pc,\rho((0,1),pc),1}} \times \inf\left(\frac{slope_{11}}{slope_{10}}, slope_{11}, 1\right) \\
& + \frac{1}{C_{g,\rho((0,0),g),0}} \times \inf\left(\frac{slope_{11}}{slope_{10} \times slope_{01}}, \frac{slope_{11}}{slope_{01}}, \frac{1}{slope_{01}}, 1\right) \\
& = \text{période}
\end{aligned}$$

6.4.2 Glissement imposé sur (1, 0)

De la même façon que précédemment, nous pouvons déterminer les valeurs possibles des délais passés dans chacun des états :

$$\begin{aligned}
- \delta_{11}^g &= \frac{1}{C_{g,\rho((1,1),g),1}} \\
- \delta_{01}^{pc} &= \frac{1}{C_{pc,\rho((0,1),pc),1}} \times \inf(slope_{11}, 1) \\
- \delta_{00}^g &= \frac{1}{C_{g,\rho((0,0),g),0}} \times \inf\left(\frac{slope_{11}}{slope_{01}}, \frac{1}{slope_{01}}, 1\right) \\
- \delta_{10}^{pc} &= \frac{1}{C_{pc,\rho((1,0),pc),0}} \times \inf\left(\frac{slope_{11} \times slope_{00}}{slope_{01}}, \frac{slope_{00}}{slope_{01}}, slope_{00}, 1\right)
\end{aligned}$$

La contrainte est donc la suivante :

$$\begin{aligned}
& \frac{1}{C_{g,\rho((1,1),g),1}} \\
& + \frac{1}{C_{pc,\rho((0,1),pc),1}} \times \inf(\text{slope}_{11}, 1) \\
& + \frac{1}{C_{g,\rho((0,0),g),0}} \times \inf\left(\frac{\text{slope}_{11}}{\text{slope}_{01}}, \frac{1}{\text{slope}_{01}}, 1\right) \\
& + \frac{1}{C_{pc,\rho((1,0),pc),0}} \times \inf\left(\frac{\text{slope}_{11} \times \text{slope}_{00}}{\text{slope}_{01}}, \frac{\text{slope}_{00}}{\text{slope}_{01}}, \text{slope}_{00}, 1\right) \\
& = \text{période}
\end{aligned}$$

6.4.3 Glissement imposé sur (1, 1)

Dans ce cas :

$$\begin{aligned}
- \delta_{01}^{pc} &= \frac{1}{C_{pc,\rho((0,1),pc),1}} \\
- \delta_{00}^g &= \frac{1}{C_{g,\rho((0,0),g),0}} \times \inf\left(\frac{1}{\text{slope}_{01}}, 1\right) \\
- \delta_{10}^{pc} &= \frac{1}{C_{pc,\rho((1,0),pc),0}} \times \inf\left(\frac{\text{slope}_{00}}{\text{slope}_{01}}, \text{slope}_{00}, 1\right) \\
- \delta_{11}^g &= \frac{1}{C_{g,\rho((1,1),g),1}} \times \inf\left(\frac{\text{slope}_{00}}{\text{slope}_{10} \times \text{slope}_{01}}, \frac{\text{slope}_{00}}{\text{slope}_{10}}, \frac{1}{\text{slope}_{10}}, 1\right)
\end{aligned}$$

Et par conséquent, la contrainte devient :

$$\begin{aligned}
& \frac{1}{C_{pc,\rho((0,1),pc),1}} \\
& + \frac{1}{C_{g,\rho((0,0),g),0}} \times \inf\left(\frac{1}{\text{slope}_{01}}, 1\right) \\
& + \frac{1}{C_{pc,\rho((1,0),pc),0}} \times \inf\left(\frac{\text{slope}_{00}}{\text{slope}_{01}}, \text{slope}_{00}, 1\right) \\
& + \frac{1}{C_{g,\rho((1,1),g),1}} \times \inf\left(\frac{\text{slope}_{00}}{\text{slope}_{10} \times \text{slope}_{01}}, \frac{\text{slope}_{00}}{\text{slope}_{10}}, \frac{1}{\text{slope}_{10}}, 1\right) \\
& = \text{période}
\end{aligned}$$

6.4.4 Glissement imposé sur (0, 1)

Dans ce cas :

$$\begin{aligned}
- \delta_{00}^g &= \frac{1}{C_{g,\rho((0,0),g),0}} \\
- \delta_{10}^{pc} &= \frac{1}{C_{pc,\rho((1,0),pc),0}} \times \inf(\text{slope}_{00}, 1) \\
- \delta_{11}^g &= \frac{1}{C_{g,\rho((1,1),g),1}} \times \inf\left(\frac{\text{slope}_{00}}{\text{slope}_{10}}, \frac{1}{\text{slope}_{10}}, 1\right)
\end{aligned}$$

$$- \delta_{01}^{pc} = \frac{1}{C_{pc,\rho((0,1),pc),1}} \times \inf \left(\frac{slope_{11} \times slope_{00}}{slope_{10}}, \frac{slope_{11}}{slope_{10}}, slope_{11}, 1 \right)$$

Et par conséquent, la contrainte devient :

$$\begin{aligned} & \frac{1}{C_{g,\rho((0,0),g),0}} \\ & + \frac{1}{C_{pc,\rho((1,0),pc),0}} \times \inf (slope_{00}, 1) \\ & + \frac{1}{C_{g,\rho((1,1),g),1}} \times \inf \left(\frac{slope_{00}}{slope_{10}}, \frac{1}{slope_{10}}, 1 \right) \\ & + \frac{1}{C_{pc,\rho((0,1),pc),1}} \times \inf \left(\frac{slope_{11} \times slope_{00}}{slope_{10}}, \frac{slope_{11}}{slope_{10}}, slope_{11}, 1 \right) \\ & = \text{période} \end{aligned}$$

6.5 Conclusion

Dans ce chapitre, nous avons présenté le cycle circadien et les mécanismes moléculaires sous-jacents, ainsi que sa capacité à s'adapter et anticiper les changements d'environnement entre le jour et la nuit. Nous avons ensuite déterminé les contraintes à satisfaire pour encoder, dans un modèle très simple (2 variables) de l'horloge, une oscillation entretenue malgré une absence de lumière. Généralement, il n'est pas possible d'obtenir un modèle différentiel présentant une oscillation entretenue avec deux variables. Notre formalisme hybride, qui met à profit une forte abstraction, permet ainsi d'expliquer intuitivement la présence d'un cycle limite. Cet atout de notre modélisation est largement expliqué par l'abstraction de tous les phénomènes plus concrets qu'offrent généralement les systèmes d'équations différentielles.

Le formalisme hybride permet, avec un raisonnement géométrique simple, d'obtenir les conditions nécessaires et suffisantes pour obtenir un cycle limite en 2 dimensions, en fonction des lieux de glissements. Ainsi, la connaissance de saturation ou de dégradation totale (qui se traduira par un glissement sur les murs extérieurs) est une information importante pour la construction des modèles surtout lorsque le système présente des oscillations.

Dans le chapitre suivant, nous allons utiliser cette propriété, ainsi que les contraintes déterminées dans ce chapitre pour construire un modèle réaliste de l'horloge circadienne.

Chapitre 7

Rafinements des modèles circadiens avec CircaPlot

Dans le chapitre précédent, nous avons vu que le cycle circadien est impliqué dans bon nombre de mécanismes physiologiques importants chez les mammifères. De plus en plus de modèles formels utilisés pour étudier diverses pathologies ont besoin de prendre en compte la chronologie quotidienne. Par conséquent, un modèle simple du cycle circadien vu comme un module serait idéal pour la construction de ces modèles. Ce chapitre construit un tel modèle qui, de manière remarquable, repose essentiellement sur deux variables (sans compter l’alternance Jour/Nuit qui agit ponctuellement sur l’horloge circadienne). Cela montre que le passage de la modélisation discrète à l’hybride fournit une précision considérable qui permet à de tout petits modèles d’être biologiquement crédibles.

Ce chapitre est dédié à la construction de ce module du cycle circadien et en particulier, à l’identification de ses célérités. Après avoir présenté le programme de simulation qui a servi à l’affinage et à la vérification du modèle simplifié de l’horloge, nous présentons deux modèles qui diffèrent par les célérités, en particulier par le signe de l’une d’elles. La première version exhibe un état stable, ce qui a pour conséquence de rendre le modèle trop robuste, alors que la seconde version remplace cet état stable par une célérité très faible, ce qui rend le modèle beaucoup plus réaliste.

7.1 CircaPlot : un programme de visualisation des trajectoires

Au cours de ma thèse, j’ai développé un programme de simulation en Python. Il m’a permis de tester des jeux de paramètres et de visualiser les traces obtenues. J’y ai intégré la lumière comme seul *zeitgeber*.

7.1.1 Motivations

Les contraintes obtenues dans la section précédente, sur la base de propriétés géométriques simples comme le théorème de Thalès, ne s’appliquent qu’aux célérités dans les conditions sans lumière. Afin d’avoir un modèle convenable, il est nécessaire que les autres célérités aient une valeur produisant un comportement réaliste en accord avec les nombreuses configurations d’alternance jour/nuit présentant un intérêt biologique. Il n’est bien sûr pas réaliste de refaire à la main les mêmes raisonnements géométriques pour chacune des conditions d’alternance, d’autant plus que les célérités mises en jeu d’une condition d’alternance à l’autre ne sont pas indépendantes. Par ailleurs, pour bon nombre de configurations d’alternance, nous n’avons pas de prédicats rigoureux qui définissent le comportement attendu.

Afin de valider ou invalider un comportement, l’expertise d’un biologiste est donc nécessaire. De nombreuses discussions avec lui permettent d’établir des classes de compor-

tements attendus pour juger les traces obtenues et donc la qualité du modèle. Pour faciliter ces discussions, un outil de visualisation dont les sorties sont les plus explicites possible s'est avéré absolument nécessaire. Cela a conduit à une procédure d'identification des paramètres par essais-erreurs. Nous allons donc soumettre le modèle à des perturbations du cycle du zeitgeber telles que les saisons, ou encore les décalages horaires, les résultats nous permettant d'ajuster les célérités du modèle pour s'approcher d'un comportement réaliste.

Le programme présenté dans cette section est donc un simulateur permettant de générer plusieurs types de représentations des traces d'une simulation.

7.1.2 Vision utilisateur

Le programme présenté ici a, par conséquent, pour objectif d'effectuer des simulations d'un réseau donné en entrée et d'exporter différents types de représentation des traces de ces simulations.

L'entrée est constituée de deux fichiers contenant :

- un réseau génétique hybride dont tous les paramètres sont fixés. Il s'agit d'une portion de programme Python isolée du reste du code qui permet de définir plus facilement le réseau avec ses célérités.
- une bibliothèque de conditions expérimentales. Il s'agit également une portion du programme Python qui définit par des variables globales les différentes conditions. Chacune d'elles contient un nom, un scénario d'oscillations jour/nuit et un état initial.

Le programme effectue une simulation sur un temps indiqué en paramètre (en heures) pour chacune des conditions choisies. À la fin de la simulation, le programme permet d'exporter :

- la période locale de chacune des oscillations afin de suivre l'évolution de cette période au cours du temps,
- une représentation au cours du temps de chacune des variables du système pendant toute la simulation, sous la forme d'une image PNG (voir Figure 7.1 gauche),
- les données des traces de chacune des variables dans un fichier CSV,
- un graphe d'états en 2 dimensions, c'est-à-dire une projection sur 2 variables de la trajectoire (Figure 7.1 droite). Le zeitgeber, quant à lui, est représenté par une alternance des couleurs de la trajectoire (bleue pour la nuit et rouge pour le jour). Le programme laisse la possibilité d'éclater la trajectoire sur plusieurs projections afin d'éviter les superpositions : typiquement un morceau de trajectoire par 24 heures, le point d'arrêt d'un morceau étant le point de départ du suivant. Chaque morceau contient un point noir représentant le début de la trajectoire dans ce morceau. Cette sortie est un fichier \LaTeX dans lequel chaque trajectoire est codée en *tikz* (un langage de dessin vectoriel interprété par \LaTeX).

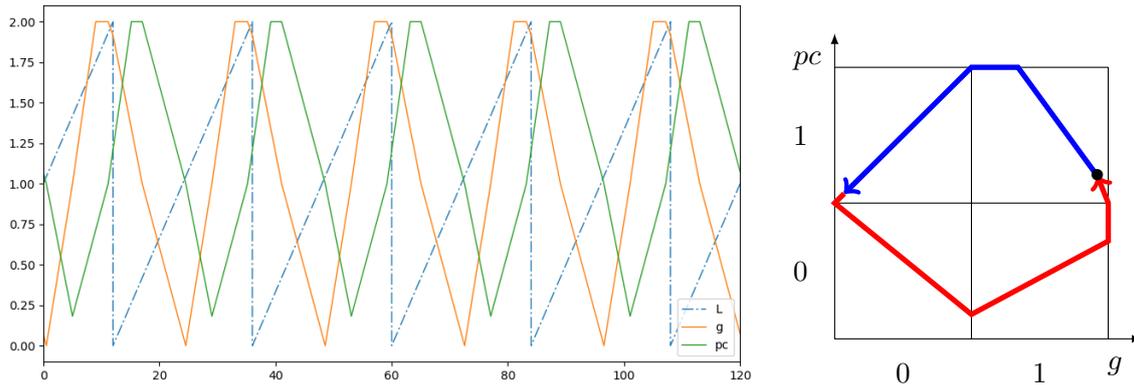


FIGURE 7.1 – Exemples de représentation des traces simulations. À gauche, l'évolution des variables au cours du temps. À droite, l'espace des phases sur 2 dimensions. Les trajectoires bleues et rouges correspondent respectivement à la nuit et au jour, et le point noir correspond à l'état initial du morceau de simulation d'une durée de 24 heures.

Chacune de ces représentations apporte une information différente dans l'étude du comportement du réseau. Les périodes successives permettent de connaître la robustesse du cycle. La représentation temporelle facilite la discussion avec les biologistes du fait de leur habitude de ce type de représentation. La représentation de l'espace des phases en 2 dimensions permet de mieux visualiser le cycle circadien afin de distinguer les pistes de modifications à apporter pour l'amélioration du modèle.

7.1.3 Structure du programme

Techniquement, le programme est écrit en Python, et, hormis les deux fichiers d'entrée, il est composé de 5 fichiers pour un total de 700 lignes :

plot Ce fichier contient principalement la fonction `plot` qui joue le rôle d'Interface Homme-Machine et permet le lancement de la simulation, ainsi que l'exportation des sorties présentées plus haut. Les arguments de cette fonction sont le réseau, la condition expérimentale, le temps de simulation et le nombre d'heures à représenter par diagramme de phase. Elle laisse également la possibilité de choisir les représentations souhaitées en sortie.

network Ce fichier contient la classe `Network` qui, comme son nom l'indique, est la classe représentant les réseaux génétiques. Chacun de ses attributs contient un des ensembles du quadruplet. Les méthodes permettent entre autres de calculer l'état hybride suivant en fonction d'un état initial donné en paramètre.

formula Ce fichier contient la classe `Formula` qui encode les formules logiques des multiplexes. Les méthodes de cette classe permettent d'analyser une chaîne de caractères et la convertir dans une structure arborescente, ainsi que de tester la validité de la formule en fonction de l'état du réseau.

state Ce fichier contient la classe `State` des états hybrides. De manière évidente, les

deux attributs qu'elle contient correspondent respectivement à la partie discrète et la partie continue de chaque état.

simulation Ce fichier contient la classe `Simulation` dont les méthodes permettent de lancer la simulation, puis d'exporter les différentes représentations des résultats vues précédemment.

Le graphe d'appel de ces différentes classes et fonctions est fourni en Figure 7.2. On voit que `plot` fait essentiellement appel à `Simulation` après avoir collecté le réseau d'entrée et les conditions expérimentales. La classe `simulation`, quant à elle, fait l'essentiel du travail en interne.

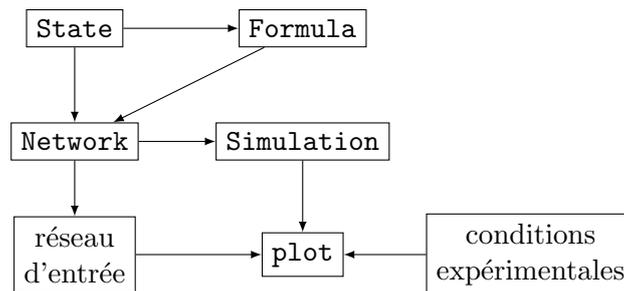


FIGURE 7.2 – Structure interne du programme de simulation

7.1.4 Protocoles d'identification des célérités

Grâce au programme `CircaPlot`, il devient aisé d'expérimenter le modèle dans de nombreuses conditions de perturbation du cycle d'alternance jour/nuit, telles que les saisons ou les décalages horaires. La visualisation offerte par `CircaPlot` facilite la démarche par essais-erreurs pilotée par un biologiste. Elle permet de plus de bien visualiser l'impact d'une célérité sur une trace globale, facilitant par là même la mise au point des célérités pour atteindre un comportement réaliste. Ainsi `CircaPlot` permet de déterminer rapidement quelles sont les célérités à modifier et comment les modifier pour corriger la trajectoire dans les différentes conditions.

En revanche, contrairement aux trajectoires discrètes du formalisme de Thomas, il est très difficile d'exprimer les trajectoires continues des modèles hybrides en logique temporelle. Parmi les difficultés, mentionnons la question des « approximations acceptables ». En effet, les parties fractionnaires étant des réels, même si un cycle dans l'espace des phases n'est pas « exactement fermé » en raison de valeurs de parties fractionnaires légèrement différentes, l'expérimentateur pourra juger que le retour au cycle limite est sensiblement atteint (donc sans nécessiter de cycles supplémentaires pour converger). C'est ce type de problèmes qui a motivé le développement de la sortie représentant l'espace des phases avec `CircaPlot`.

Du point de vue biologique, un critère majeur pour identifier les célérités qui restent à

déterminer est que la phase lumineuse impose au système d'avoir une période de 24 heures. Nous savons également que, jusqu'à un certain point, changer la proportion entre la phase lumineuse et la phase d'obscurité ne change pas la période. « Jusqu'à un certain point », car, dans ces situations extrêmes telles que dans une condition d'obscurité constante, l'horloge reprend sa période endogène, alors que dans une condition de lumière constante, elle adopte une période plus grande [65]. Ce sont donc successivement ces différentes conditions d'alternances jour/nuit que nous avons exploitées, dans un premier temps, pour identifier les célérités manquantes, et dans un second temps (sections suivantes), pour valider ou invalider le modèle obtenu.

L'identification des célérités manquantes a été faite en partant de délais (approximatifs) trouvés dans la littérature (en particulier dans [21]). Nous avons ensuite modifié manuellement les célérités de jour afin d'observer des comportements biologiquement réalistes, typiquement le respect des horaires d'acrophase¹ et de batyphase² de nos variables, ainsi que la production d'une période correcte. Par exemple, l'une des dernières étapes de ce processus d'identification des célérités manquantes a été le changement de la célérité $C_{g,\{m_L\},0}$. La figure 7.3 présente la différence qu'a entraîné le changement de cette célérité sur la dynamique. À l'origine, la valeur était de $-\frac{1}{7.5}$, mais la période d'oscillation était supérieure à 24 heures (figure de droite). Nous avons alors ralenti la décroissance de g sur la section (1). L'effet de cette modification est la remontée de la batyphase (2) qui permet de raccourcir le cycle et donc de diminuer la période.

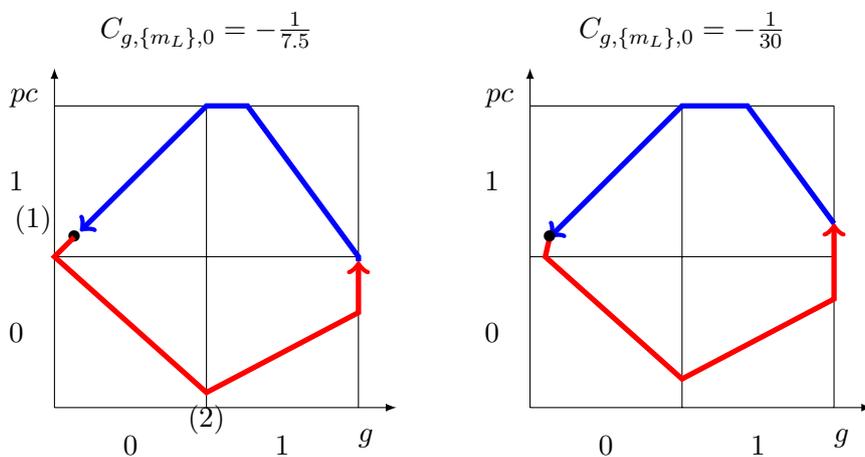


FIGURE 7.3 – Effet du changement de la valeur d'un paramètre sur la dynamique du modèle. À gauche, avant modification (période supérieure à 24 heures) et à droite après (période de 24 heures).

1. Maximum d'une variable qui est soumise à un cycle biologique.
2. Minimum d'une variable qui est soumise à un cycle biologique.

7.2 Premier modèle : avec état stable

Avant d'appliquer notre formalisme à un modèle du noyau suprachiasmatique, nous avons simplement transcrit sous forme de célérités le modèle proposé dans [21], qui est un modèle où la lumière agit sur le complexe PER-TIM dans le noyau de la cellule. Il s'agit en fait d'un modèle dédié à la drosophile (chez qui TIM joue le rôle que joue CRY chez le mammifère). Nous le présentons ici parce qu'il a été le premier à réduire l'horloge circadienne à 2 variables et à fournir une intuition simple à propos de l'effet des changements de proportions jour/nuit.

Ce premier essai ayant été concluant pour l'étude de la résistance aux saisons, nous avons par la suite adapté cette vision à 2 variables aux mammifères. Ceci avec deux objectifs principaux : préserver les intuitions simples offertes par le précédent modèle, et mieux représenter l'action de la lumière sur l'horloge circadienne dans le cas des mammifères. Afin d'éviter les répétitions, nous ne présenterons le modèle complet avec célérités que pour le modèle mammifère.

7.2.1 Modèle discret pour la drosophile

Les différences entre l'horloge de la drosophile et celle des mammifères sont très peu nombreuses au niveau d'abstraction dans lequel on se place. Le gène *Cry* est remplacé par un homologue³, le gène *Tim*. Les protéines CRY sont néanmoins présentes chez la drosophile et jouent un rôle de régulateur, car elles sont activées par la lumière, ce qui leur permet d'inhiber PER-TIM *via* la dissociation du complexe. De ce fait, comme montrée dans la figure 7.4, la régulation du zeitgeber s'effectue négativement sur la variable *pt* (représentant le complexe PER-TIM).

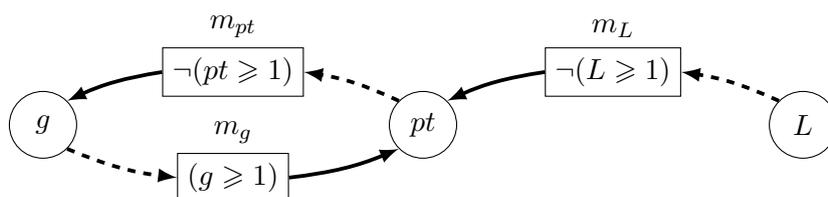


FIGURE 7.4 – Graphe d'interaction de l'horloge moléculaire circadienne de la drosophile

L'approche présentée dans cet article [21] est basée sur le formalisme discret de modélisation de Thomas, auquel a été ajouté un ensemble de délais pour chacun des états discrets. Les paramètres choisis font apparaître un état qui est stable en phase de jour, mais qui ne l'est plus en phase nocturne. C'est cet état stable qui permet une robustesse du modèle (voir Figure 7.5).

L'intérêt de cette approche discrète est qu'elle montre clairement que, lorsque la lumière

3. homologues : gènes différents ayant la même fonction

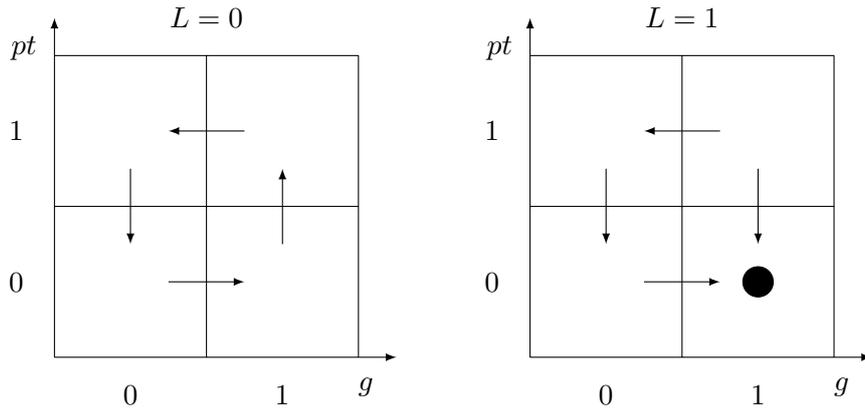


FIGURE 7.5 – Dynamique du réseau de la figure 7.4. À gauche le comportement du réseau en phase nocturne, et à droite en phase diurne.

est présente, pour une durée imposée suffisamment longue, le système entre dans un état stable ($g = 1$ et $pt = 0$ en figure de droite). De ce fait, en période d'été, l'allongement de la journée ne perturbe pas le système puisqu'il attend dans l'état stable et que le cycle repartira de l'état $(1,0)$ lorsque la nuit arrivera. Comme démontré dans [21], la même idée explique très clairement que l'on supporte bien un décalage horaire qui allonge la journée (voyage vers l'ouest). En revanche, un décalage horaire vers l'est qui écourte la nuit, peut faire passer le système de la configuration de gauche à la configuration de droite pendant que la transition sortante de $(1,0)$ a lieu. Ceci engendre une inversion de l'évolution de pt qui peut expliquer que l'on supporte moins bien un décalage horaire vers l'est.

Pour étudier ces questions plus spécifiquement chez le mammifère, nous ne pouvons pas utiliser ce modèle tel quel, puisque la régulation par la lumière doit se faire *via* les gènes *Per* et par conséquent, comme déjà vu, l'inhibition de pt par la lumière doit être remplacée par une activation de g . Évidemment, il faut également remplacer la variable pt qui représente PER-TIM par pc qui représente PER-CRY. C'est ce que nous faisons dans la sous-section suivante.

7.2.2 Modèle mammifère avec état stable

Dans cette section, nous considérons, par conséquent, le modèle de la figure 6.2 en page 99. Comme déjà mentionné, nous sommes partis de délais issus de la littérature (essentiellement [21, 42]) pour construire un premier jeu de valeurs pour les différentes célérités; nous avons ensuite appliqué la méthode décrite dans la section 7.1.4 pour adapter plus finement les comportements du modèle à des observations faites chez le mammifère.

- Selon les individus, la période intrinsèque (sans zeitgeber) est entre 24 et 26h [47]. Ici, nous avons fait le choix d'une période endogène (condition d'obscurité constante) de l'horloge de 26 heures, pour mieux distinguer l'effet de la lumière.
- Nous avons ensuite déterminé les états dans lesquels s'effectuent les glissements en

obscurité ($L = 0$) : sur la borne supérieure de g (état $(1, 0)$) et de pc (état $(1, 1)$). Ces deux glissements représentent respectivement l'accumulation des protéines PER et CRY dans le cytoplasme, et la saturation de complexes PER-CRY dans le noyau [21].

Les célérités obtenues sont indiquées dans la table 7.1.

Célérités de g	Célérités de pc
$C_{g,\emptyset,0} = -\frac{1}{7,5}$	$C_{pc,\emptyset,0} = -\frac{1}{5}$
$C_{g,\emptyset,1} = -\frac{1}{5,5}$	$C_{pc,\emptyset,1} = -\frac{1}{7,5}$
$C_{g,\{m_{pc}\},0} = \frac{1}{5,5}$	$C_{pc,\{m_g\},0} = \frac{1}{7,5}$
$C_{g,\{m_{pc}\},1} = \frac{1}{4}$	$C_{pc,\{m_g\},1} = \frac{1}{4}$
$C_{g,\{m_L\},0} = -\frac{1}{30}$	
$C_{g,\{m_L\},1} = \frac{1}{5,5}$	
$C_{g,\{m_{pc},m_L\},0} = \frac{1}{4,5}$	
$C_{g,\{m_{pc},m_L\},1} = \frac{1}{4}$	

TABLE 7.1 – Valeurs de célérités pour le modèle

Durant cette démarche, nous avons voulu mimer le phénomène de l'état stable décrit précédemment sur la drosophile. Cela a conduit à un jeu de célérités dans lequel, pour une variable donnée et un ensemble de ressources donné, le signe des célérités ne dépend pas de l'état de la variable. Par exemple, $C_{pc,\{m_g\},0}$ et $C_{pc,\{m_g\},1}$ sont toutes les deux positives. Il y a une seule exception à ce principe : $C_{g,\{m_L\},0}$ est négative et $C_{g,\{m_L\},1}$ est positive. Les deux célérités de l'état $(1, 1)$ sont alors positives et le coin supérieur droit devient stable en présence de lumière (Figure 7.6).

Une remarque importante : pour constituer ce modèle en restant fidèle à la philosophie du modèle discret décrit en section 7.2.1, nous avons imposé un état stable unique situé dans le coin extrême de l'état $(1,1)$. Pour ce faire, il est nécessaire de que la célérité en g soit positive dans cet état. Cependant, dans l'état $(0,1)$ la célérité en g doit être négative pour permettre la reprise du cycle. *In fine*, cela nous a conduits aux valeurs $C_{g,\{m_L\},0} = -\frac{1}{30}$ et $C_{g,\{m_L\},1} = \frac{1}{5,5}$. Ces valeurs ne sont pas compatibles avec la définition 13, car il n'est pas possible d'avoir des célérités divergentes sans changer de ressources. La question qui s'est alors posée à nous, était de savoir s'il fallait relaxer la définition 13 afin d'autoriser de tels états stables « en coin ». Nous répondrons à cette question dans la section 7.2.6.

Dans la suite de cette section, nous allons voir que ce modèle s'adapte « trop bien » aux perturbations que nous lui imposons : il faut des perturbations très importantes pour pouvoir observer des comportements problématiques biologiquement. Cette robustesse extrême justifiera alors la conception d'un nouveau modèle présenté dans la section 7.3.

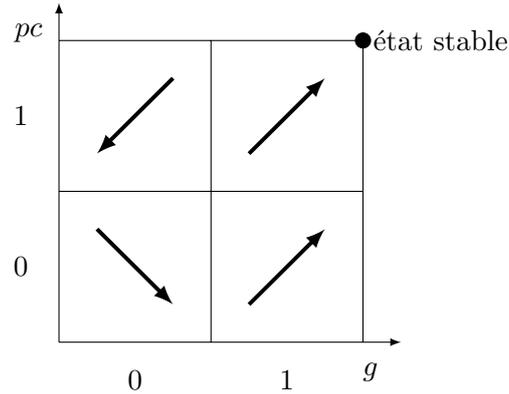
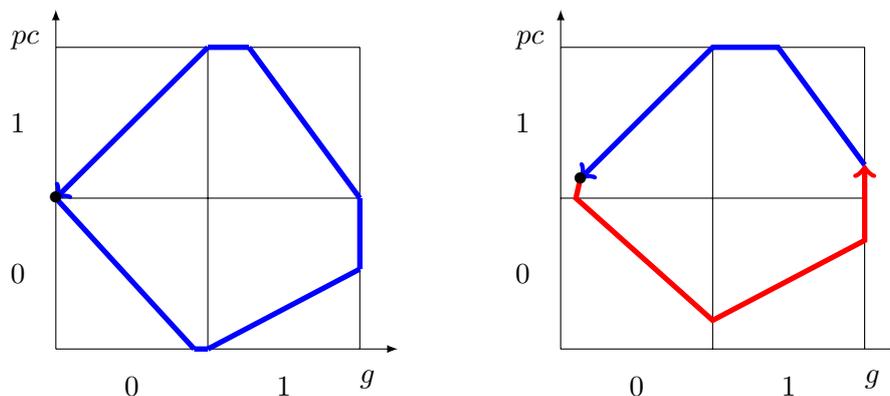


FIGURE 7.6 – Localisation de l'état stable dans la projection du graphe d'états sur les variables g et pc avec $L = 1$

7.2.3 Test du modèle en obscurité constante et en équinoxe

Les premières conditions testées sont les conditions d'obscurité constante et d'équinoxe (12 heures de lumière et 12 heures d'obscurité) :

- Les contraintes dans la section 6.2.2 étant déterminées pour obtenir le cycle limite endogène de l'horloge, la condition d'obscurité constante donne les résultats attendus, dans la représentation de gauche ci-dessous. Cette représentation correspond à une trajectoire dans l'espace des états de 26 heures.
- La trajectoire de droite correspond à 24 heures avec 12 heures de nuit puis 12 heures de jour. On remarque que la lumière raccourcit la trajectoire en bas de l'espace (partie rouge). On peut remarquer également que la trajectoire glisse le long de la borne supérieure de g dans l'état $(1, 1)$. Ce comportement correspond à l'approche de l'état stable du plan $L = 1$. Dès l'obscurité revenue ($L = 0$), la trajectoire bleue s'écarte à nouveau du bord droit pour reprendre le cycle.



7.2.4 Test de résistance du modèle aux saisons

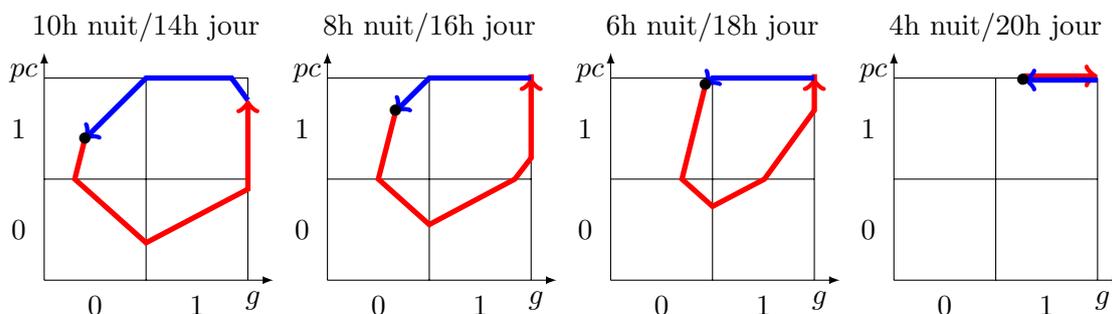
Les phénomènes les plus courants auxquels doit faire face l'horloge circadienne sont les saisons. En effet, elle doit s'adapter aux changements de proportion entre la phase de jour et la phase de nuit.

Dans le cas de l'été, nous avons simulé avec 4 conditions différentes :

- 10 heures d'obscurité, 14 heures de lumière
- 8 heures d'obscurité, 16 heures de lumière
- 6 heures d'obscurité, 18 heures de lumière
- 4 heures d'obscurité, 20 heures de lumière

Les figures ci-dessous représentent les trajectoires de ces 4 comportements (de gauche à droite). On voit que le raccourcissement de la phase d'obscurité (trajectoire bleue) a pour effet d'écourter la trajectoire rouge. Avec une durée d'obscurité inférieure à 6h, g n'est plus réprimé suffisamment longtemps et la trajectoire rouge ne permet plus à pc d'atteindre son seuil. Cette variable est donc bloquée à sa valeur 1. L'horloge n'est alors plus fonctionnelle : il n'y a plus d'oscillations qualitatives.

Dans ces 4 conditions, malgré les différences notables de trajectoires, la période est toujours de 24 heures. Cela est dû à la présence de l'état d'attente (point noir en haut à droite). Même si la trajectoire ne sort pas de l'état $(1, 1)$ (comme dans la condition 4/20), le système évolue un peu puis revient à l'état d'attente, et ne repart qu'à l'apparition de la nouvelle phase de nuit.

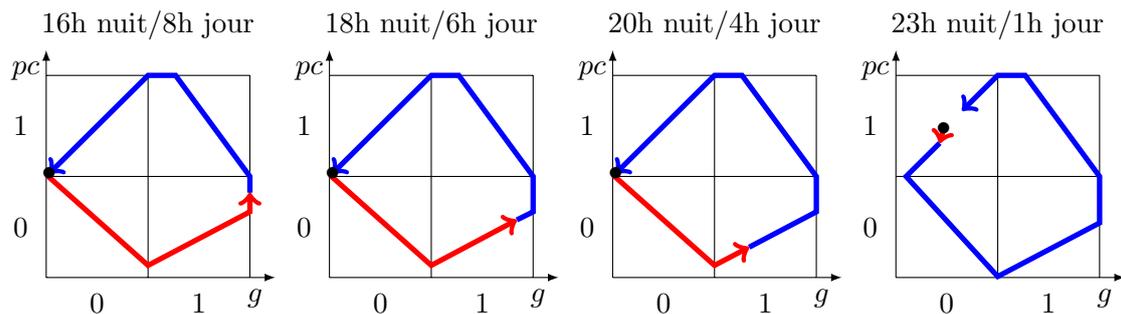


Regardons maintenant le comportement du réseau avec une phase de nuit plus importante (hiver). Nous avons simulé des conditions inverses à celles de l'été, autrement dit :

- 16 heures d'obscurité, 8 heures de lumière
- 18 heures d'obscurité, 6 heures de lumière
- 20 heures d'obscurité, 4 heures de lumière
- 23 heures d'obscurité, 1 heure de lumière

Les figures ci-dessous représentent les comportements du réseau dans ces 4 conditions. Cette fois, il n'y a pas de différences qualitatives du comportement entre elles (elles passent

par les mêmes états discrets). Dans les 3 premières conditions (16/8, 18/6 et 20/4), nous pouvons simplement distinguer la diminution de la phase de lumière (trajectoire rouge). La période ne change pas non plus. Dans la dernière condition, nous pouvons voir que le cycle n'est pas fermé, donc la période est légèrement supérieure à 24 heures : dans cette condition, la phase de lumière n'est plus assez grande pour synchroniser la période à 24 heures.



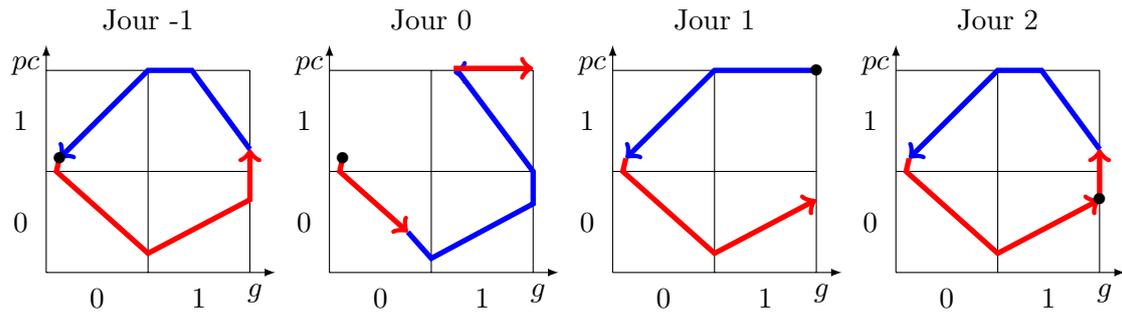
7.2.5 Test de résistance du modèle aux décalages horaires

Les expérimentations autour des décalages horaires sont un bon moyen de tester la robustesse d'un modèle circadien. Ici, nous avons testé un décalage horaire de 8h. Les 4 conditions sont les suivantes :

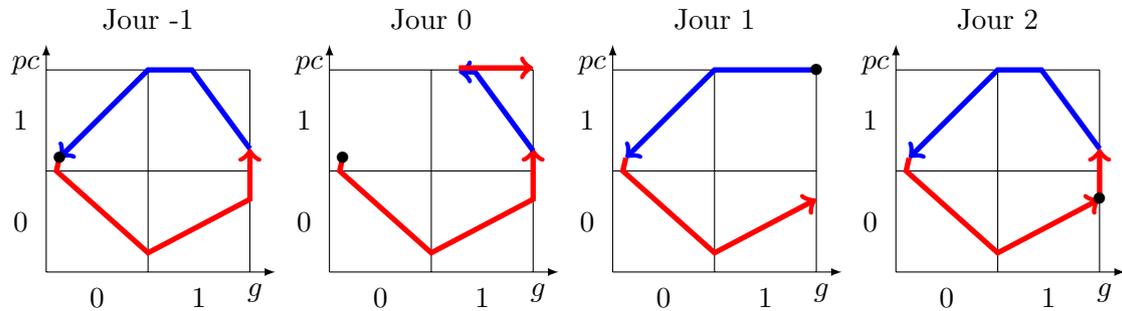
- Raccourcissement de la phase lumineuse (vers l'est le jour)
- Raccourcissement de la phase d'obscurité (vers l'est la nuit)
- Allongement de la phase lumineuse (vers l'ouest le jour)
- Allongement de la phase d'obscurité (vers l'ouest la nuit)

Dans chacune des 4 figures ci-dessous, le premier cycle (jour -1) est un cycle d'équinoxe « normal » afin de démarrer la simulation avec des conditions standard. La perturbation du décalage horaire a lieu lors des 24 heures suivantes, puis les tranches de 24 heures d'après (notées jour 1 et jour 2) sont soumises à une alternance jour/nuit d'équinoxe. Nous pouvons alors y observer le comportement de resynchronisation du système dans chacune des 4 figures ci-dessous.

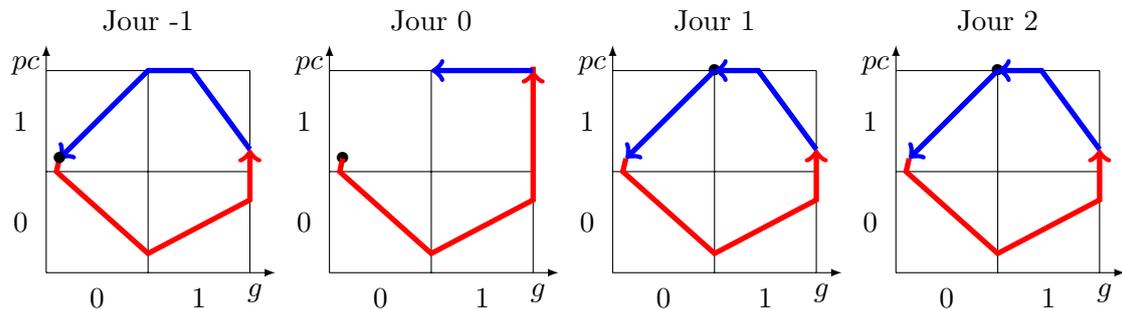
La première figure ci-dessous contient les résultats de la première condition (vers l'est le jour). Nous pouvons voir que le raccourcissement de la phase de jour fait que le cycle n'est pas fermé au jour 0, car le décalage de la phase lumineuse provoque un rabat de la trajectoire vers l'état stable. Le système attend alors le début du jour 1 pour reprendre le cycle standard. Au jour 2, le cycle est à nouveau visible.



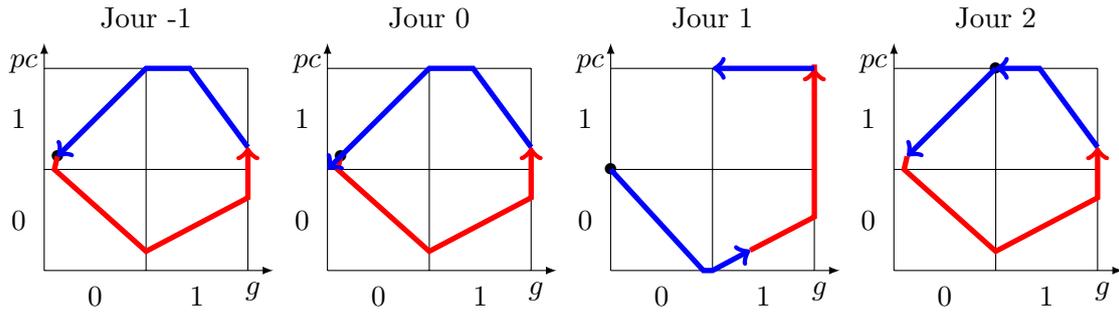
Dans la condition vers l'est la nuit, le comportement est identique (figure ci-dessous). La seule différence réside essentiellement dans la position des phases de jour et de nuit durant le jour 0.



La figure suivante représente le comportement du système lors d'un voyage vers l'ouest durant le jour. On remarque que la perturbation est moins importante que dans les deux conditions de voyage vers l'est. En effet, la perturbation consiste à créer simplement, le jour même, une attente sur l'état stable. Cela crée un cycle plus long lors du décalage (jour 0). Dès le jour 1, on peut voir la trajectoire qui a rejoint le cycle standard.



Enfin, la figure suivante représente le comportement du système lors d'un voyage vers l'ouest durant la nuit. Ce jour-là, la trajectoire bleue est rallongée de 8 heures. Durant les 24 heures suivantes (jour 1) la trajectoire est bloquée sur l'état stable de la même façon que dans la condition précédente, mais avec 1 jour de décalage. Lors du jour 2, le système a retrouvé sa trajectoire standard.



7.2.6 Bilan

En soumettant le système à une différence de proportions de phases jour/nuit (variation saisonnière), le système s'avère très robuste à une diminution du temps d'exposition à la lumière, mais très fragile à une augmentation.

Dans le cas des décalages horaires, du point de vue qualitatif, la perturbation pour un décalage vers l'est est plus importante que pour un décalage vers l'ouest. Cependant, dans tous les cas, la resynchronisation est extrêmement rapide, ce qui, du point de vue quantitatif, n'est absolument pas réaliste. Ce défaut du modèle est essentiellement dû à la violence avec laquelle l'état stable réagit aux perturbations en capturant les trajectoires.

Dans tous les cas, c'est la présence de l'état stable dans le coin de l'état (1,1) qui rend le comportement du système caricatural. Il semble donc important de prendre en compte le fait que, en réalité, les mammifères ne restent pas indéfiniment dans cet état stable, même en condition de jour constant. Dans la section suivante, nous allons donc remplacer l'état stable par une célérité très lente qui rend l'état (1,1) « presque stable » sur le court terme. L'idée est qu'une vitesse très lente dans l'état (1,1) permet quand même aux trajectoires de finir par s'échapper de cet état pour continuer d'évoluer.

De ce bilan, il ressort que l'idée d'un état stable, qui était pourtant judicieuse dans le cadre discret, ne l'est plus dans notre cadre hybride. Rappelons que pour obtenir cet état stable nous avons dû déroger à la définition 13 à propos du signe des célérités. Comme nous allons le voir, le modèle que nous présentons dans la section suivante suit la définition tout en ayant un comportement beaucoup plus crédible que le modèle avec état stable. C'est pour cette raison que nous n'avons pas relaxé les contraintes de la définition 13, et c'est important, car cela permet de préserver la compatibilité des signes des célérités avec la théorie discrète de Thomas.

7.3 Second modèle : sans état stable

Nous avons conçu le modèle précédent à partir d'un modèle de Thomas dans lequel il y avait une différence de comportement discret selon la présence ou l'absence de lumière, la lumière induisant un état stable. Cependant, avec le formalisme hybride, nous pouvons définir un modèle beaucoup plus subtil en rendant un état « presque stable ». En effet, la vision discrète induit le changement de signe de certaines célérités en fonction de la lumière, alors que le pouvoir d'expression offert par l'approche hybride permet de simplement changer des valeurs de célérité. Concrètement, la différence entre le modèle présenté dans cette section et celui de la section précédente est essentiellement le signe de la célérité $C_{g,\{m_L\},1}$ qui devient négative, comme son équivalent sans lumière $C_{g,\emptyset,1}$, mais avec une valeur absolue très petite. L'équivalent discret de ce nouveau modèle n'aurait, par contre, aucun intérêt, car les comportements avec et sans lumière seraient rigoureusement les mêmes : les signes des célérités étant partout ceux correspondant à l'absence de lumière, l'état stable induit par le jour constant n'existe plus.

La valeur de $C_{g,\{m_L\},1}$ ayant une valeur très faible permet de simuler « l'état d'attente » sans pour autant avoir la robustesse excessive du modèle précédent. La table 7.2 contient l'ensemble des paramètres de ce nouveau modèle.

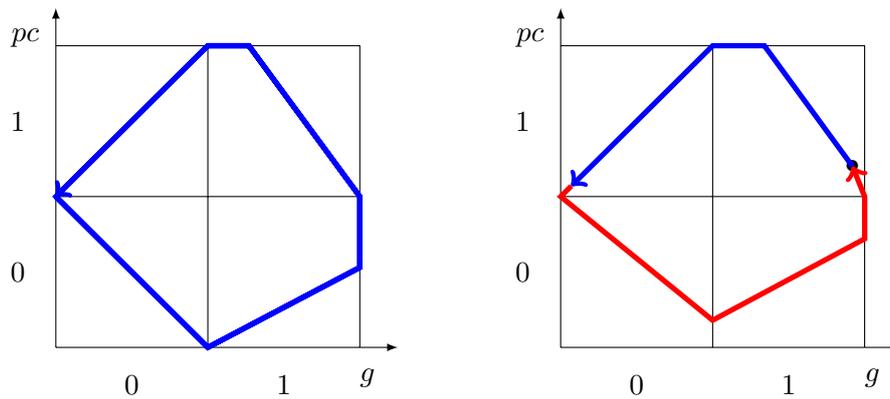
Célérités de g	Célérités de pc
$C_{g,\emptyset,0} = -\frac{1}{7,5}$	$C_{pc,\emptyset,0} = -\frac{1}{5,5}$
$C_{g,\emptyset,1} = -\frac{1}{5,5}$	$C_{pc,\emptyset,1} = -\frac{1}{7,5}$
$C_{g,\{m_{pc}\},0} = \frac{1}{5,5}$	$C_{pc,\{m_{pc}\},0} = \frac{1}{7,5}$
$C_{g,\{m_{pc}\},1} = \frac{1}{4}$	$C_{pc,\{m_{pc}\},1} = \frac{1}{4}$
$C_{g,\{m_L\},0} = -\frac{1}{7,5}$	
$C_{g,\{m_L\},1} = -\frac{1}{10}$	
$C_{g,\{m_{pc},m_L\},0} = \frac{1}{4,5}$	
$C_{g,\{m_{pc},m_L\},1} = \frac{1}{4}$	

TABLE 7.2 – Valeurs de célérités pour le modèle sans état stable

Pour identifier les célérités de cette table, nous sommes partis des délais du modèle avec état stable pour construire un premier jeu de valeurs pour les différentes célérités ; nous avons ensuite appliqué la méthode décrite dans la section 7.1.4. Outre la célérité $C_{g,\{m_L\},1}$ qui devient négative comme mentionné plus haut, nous avons été amenés à modifier $C_{g,\{m_L\},0}$. Suivant la méthode de la section 7.1.4 la valeur $-\frac{1}{10}$ de $C_{g,\{m_L\},1}$ a été calibrée pour que les délais de récupération de phase deviennent réalistes, ce que ne permettait pas le modèle avec état stable. Il est alors apparu que la célérité $C_{g,\{m_L\},0}$ de $-\frac{1}{30}$ induisait des retards en désaccord avec les connaissances biologiques, et la méthode nous a conduits à augmenter cette vitesse jusqu'à $-\frac{1}{7,5}$. De même, de manière marginale, nous avons été amenés à modifier très légèrement $C_{pc,\emptyset,0}$. Nous reprenons ci-dessous les cas de figure qui nous ont permis de mener à bien l'identification de ces paramètres ; ce sont les mêmes conditions d'alternance que dans le modèle précédent.

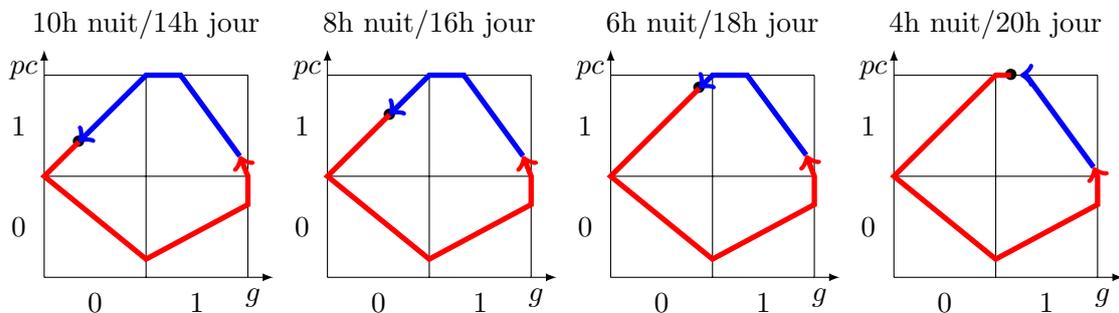
7.3.1 Test du modèle en obscurité constante et en équinoxe

La figure ci-dessous à gauche contient la trajectoire du système pendant 26 heures dans la condition d'obscurité constante, et à droite, la trajectoire pendant 24 heures dans la condition standard d'oscillation, avec 12 heures de nuit et 12 heures de jour. Pour cette dernière condition, dans l'état $(1, 1)$, la trajectoire ne reste plus contre le mur extérieur de g comme dans le modèle précédent, mais s'en éloigne lentement (dernière portion de la trajectoire rouge). Au début de la phase d'obscurité, la dégradation de pc accélère (les deux premières portions de la trajectoire bleue).



7.3.2 Test de résistance du modèle aux saisons

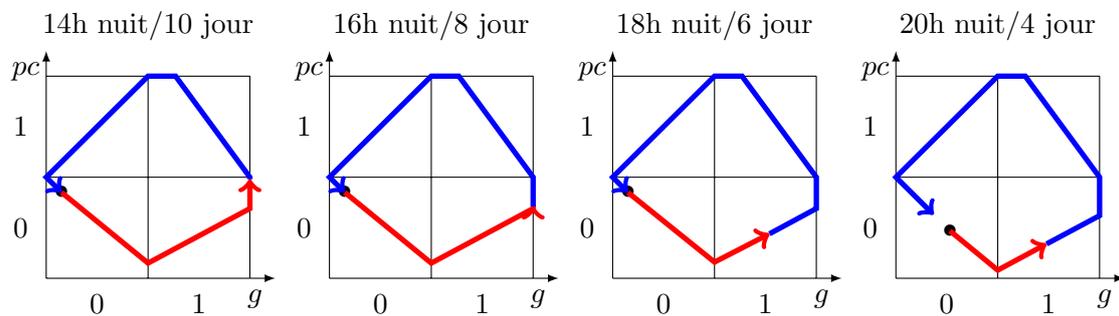
Dans le cas d'allongement de la phase lumineuse (été), les résultats obtenus sont présentés dans la figure ci-dessous. Le système est robuste pour des phases de lumière inférieures à 18 heures. Cependant, dans la simulation avec une phase lumineuse de 20 heures, la trajectoire n'arrive plus à fermer le cycle dans les 24 heures. Dans cette situation, la période est plus grande que 24 heures. La robustesse de ce nouveau modèle est donc moins importante que celle du modèle précédent comme souhaité, ce qui est plus réaliste.



Nous constatons dans la figure précédente que le point de passage de la nuit au jour change de phase en fonction des conditions. En revanche, la phase reste stable dans les 3 premiers

cas, ce qui suggère des conditions « supportables », ce qui n'est pas le cas avec seulement 4 heures de nuit où la phase se décale au fur et à mesure des cycles.

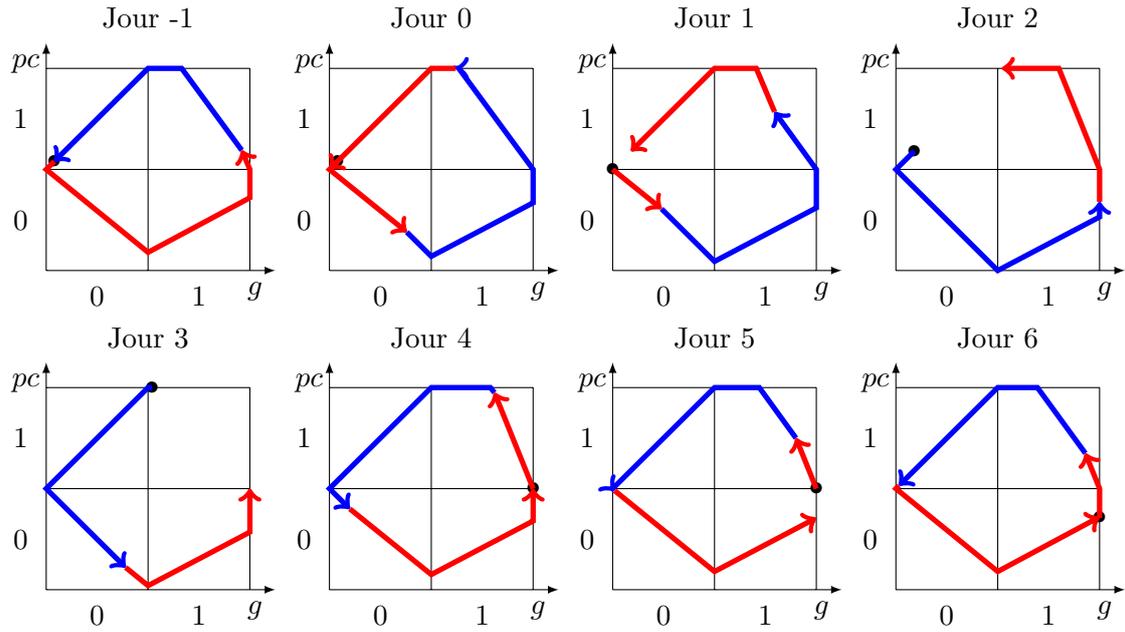
Dans le cas de raccourcissement de la phase lumineuse (hiver), le modèle a été soumis à des phases de jour de 10 heures à 4 heures. Les résultats sont visualisés dans la figure ci-dessous. De la même manière que pour l'allongement de la phase de jour, le système est robuste pour une phase supérieure à 6 heures. En dessous de ce délai, le temps d'exposition à la lumière n'est plus suffisant pour synchroniser l'horloge à 24 heures. La période tend alors vers 26 heures, période endogène de l'horloge, ce qui est réaliste.



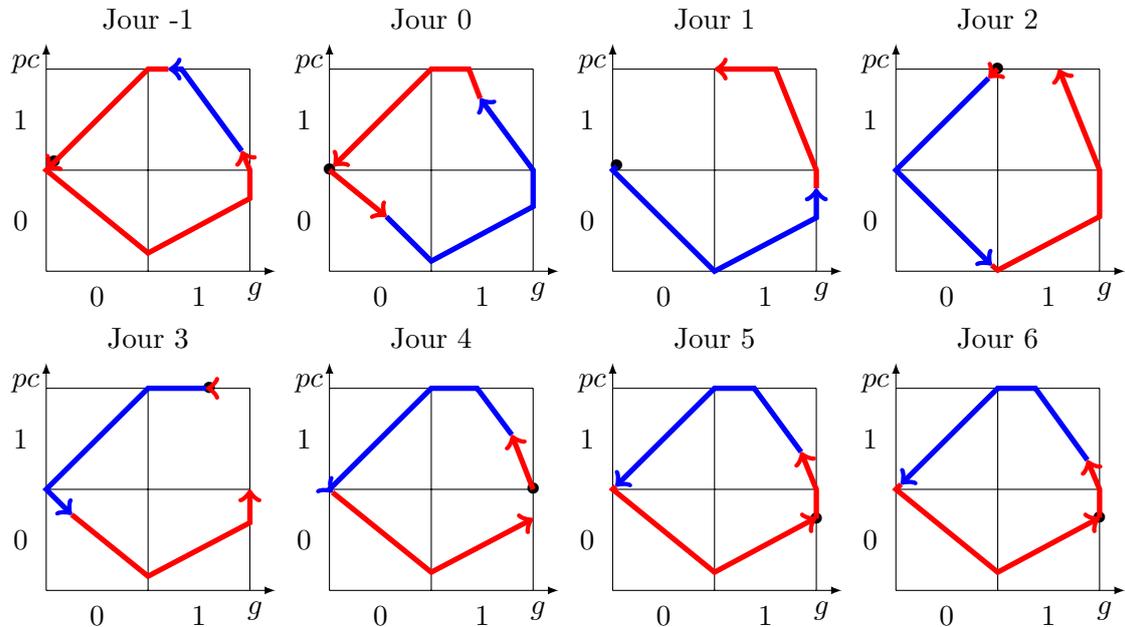
Là encore, nous constatons que le point de passage de la nuit au jour change de phase en fonction des conditions. En revanche, la phase reste stable dans les 3 premiers cas, ce qui suggère des conditions « supportables », ce qui n'est pas le cas avec seulement 4 heures de jour où la phase se décale au fur et à mesure des cycles.

7.3.3 Test de résistance du modèle aux décalages horaires

La figure ci-dessous représente l'évolution de la trajectoire après un raccourcissement de la phase lumineuse de 8 heures pendant un seul cycle (le jour 0 ne bénéficie que de 4 heures de jour). Cela correspond à un décalage horaire de 8 heures vers l'est pendant la phase de jour. Le système met alors 6 jours pour retrouver la phase initiale (jour 6). Nous pouvons voir que la trajectoire correspondant au jour (portion rouge de la trajectoire) « pivote » autour de l'intersection des seuils, et finit par retrouver sa position initiale (jour -1). Notre modèle sans état stable, une fois les célérités calibrées, permet donc de retranscrire le rephasage progressif observé couramment à l'issue d'un décalage horaire.

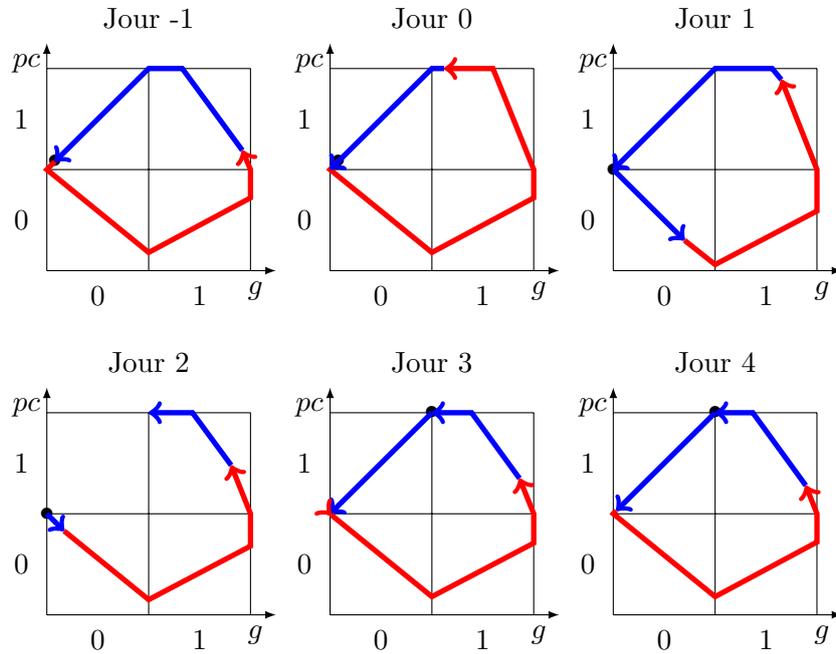


De manière similaire, nous avons effectué une simulation plaçant la réduction de 8 heures sur la phase d'obscurité (figure ci-dessous). Cela correspond à un décalage horaire, toujours vers l'est, de 8 heures pendant la phase de nuit. Le résultat est similaire, si ce n'est une réduction d'un jour sur le rephasage (5 jours suffisent, le cycle est fermé au jour 5). Ceci est cohérent avec le vécu majoritaire : un décalage horaire vers l'est pendant la phase lumineuse est plus perturbant pour l'horloge qu'un décalage pendant la nuit.

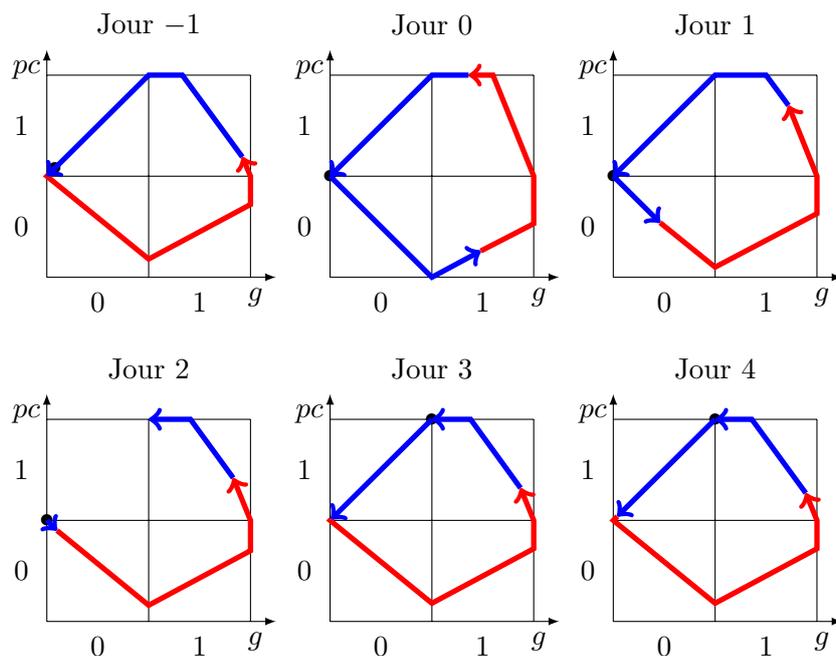


La figure suivante montre l'évolution du système lors d'un allongement de la phase lumineuse de 8 heures sur un seul jour (jour 0). À l'issue du retard de phase que le décalage

horaire provoque, le système est plus rapide à retrouver sa phase d'origine qu'avec une avance de phase. En effet, il lui suffit de 4 jours pour retrouver sa phase contre les 6 que nous avons pour un décalage horaire vers l'est.



Enfin, la dernière figure représente la dynamique du modèle dans la dernière condition : un décalage horaire de 8 heures vers l'ouest durant la nuit. La différence avec la simulation précédente est très faible puisque, dès le jour 1, les décalages de phases sont les mêmes que dans le cas précédent. *A fortiori*, 4 jours sont également nécessaires pour le rephasage.



D'après la faculté du modèle à ajuster sa phase, les décalages horaires devraient être plus précis que dans le modèle précédent. En effet, si le rephasage prend plus de temps avec une avance de phase qu'avec un retard, un décalage horaire vers l'est devrait être plus perturbant que vers l'ouest.

Pour conclure, ce second modèle est beaucoup plus proche de comportements réels que le modèle précédent. La transformation d'un état stable en état d'attente reflète beaucoup plus fidèlement la robustesse du système aux perturbations des cycles lumineux (saisons, décalages horaires). Biologiquement, un organisme se remet toujours plus difficilement d'une avance de phase et ce sont les comportements que nous retrouvons dans les simulations avec le modèle sans état stable.

7.4 Conclusion

Le modèle avec état stable est fidèle à son équivalent discret selon la théorie de Thomas, où l'état stable permet au système d'attendre l'arrivée de la nuit. Cet état le rend cependant excessivement robuste à des conditions de photopériodes extrêmes. Rappelons également qu'il impose une entorse à la définition 13. De ce fait, nous avons abandonné ce modèle hybride avec état stable. Nous avons ainsi décidé de remplacer cet état stable par une célérité lente permettant au système d'attendre, mais sans s'arrêter. En revanche, le réseau équivalent à ce modèle dans le formalisme discret de Thomas n'a aucun intérêt, car les signes des célérités engendrent des comportements qualitatifs identiques de jour comme de nuit.

Au vu de la simplicité du modèle hybride sans état stable et du comportement qui en résulte, ce modèle s'avère très satisfaisant. Avec seulement 3 variables en comptant la lumière, ce modèle de l'horloge circadienne peut être utilisé comme « module » dans les couplages avec d'autres systèmes, sur lesquels l'influence du cycle circadien est fondamentale [9]. Plus généralement, on constate que la richesse apportée par le remplacement des paramètres discrets de Thomas par des célérités réelles permet d'atteindre des comportements temporels complexes tout en préservant la simplicité des modèles.

Chapitre 8

Prolongements du formalisme hybride

Le formalisme hybride introduit dans le chapitre 5 permet d'encoder des informations temporelles réelles observées expérimentalement. Cependant, la technicité de ce cadre de modélisation constitue un frein pour pouvoir développer rapidement de nouveaux modèles avec les biologistes. La plateforme de simulation HyMBioNet [23] développée dans l'équipe s'appuie le cadre de modélisation développé dans cette thèse et permet de construire un premier modèle très rapidement, tout en rendant possible la visualisation des traces obtenues. Ceci facilite la validation/réfutation du modèle par les experts du domaine.

De plus, l'identification des paramètres étant l'une des limitations de ce cadre de modélisation, l'équipe a aussi développé, à partir du formalisme du chapitre 5, une méthode de construction de contraintes sur les paramètres, afin que le modèle puisse exhiber une trace spécifiée. Nous montrons ici comment cette méthode a permis de construire un modèle pertinent du cycle cellulaire [11].

8.1 plateforme de simulation HyMBioNet

Pour construire avec les biologistes un modèle hybride, il est utile d'avoir un simulateur dont l'interface est la plus intuitive possible. De plus, si la portabilité du simulateur est importante, il pourra être installé et utilisé directement par les experts. C'est dans ce contexte qu'en collaboration avec Gilles Enée, nous avons développé la plateforme HyMBioNet (Hybrid Models of Biological Networks).

Cette plateforme se décompose en 2 parties :

1. l'interface permet à l'utilisateur de décrire le réseau de régulation. Cette partie est développée sur un schéma de page web, en utilisant le langage PHP. L'avantage de ce choix réside dans la portabilité de la solution, puisqu'étant hébergé sur le site du laboratoire, une simple connexion internet permet d'y accéder. Cette interface crée un fichier de description du simulateur.
2. l'interface de simulation récupère le fichier de description créé à l'étape précédente et utilise l'environnement de programmation de NetLogo [77] qui propose une interface graphique intuitive.

Pour illustrer l'utilisation de cette plateforme, nous reprenons le réseau du cycle circadien utilisé dans le chapitre 6 (voir Figure 8.1).

8.1.1 Interface de description d'un réseau hybride de régulation génétique

Simuler un réseau hybride nécessite la description de l'ensemble de variables, des zeitgebers, des multiplexes, ainsi que des arcs entre multiplexes et variables. La figure 8.2 montre l'interface de description d'un tel réseau. Les deux parties « VAR » et « REG » contiennent

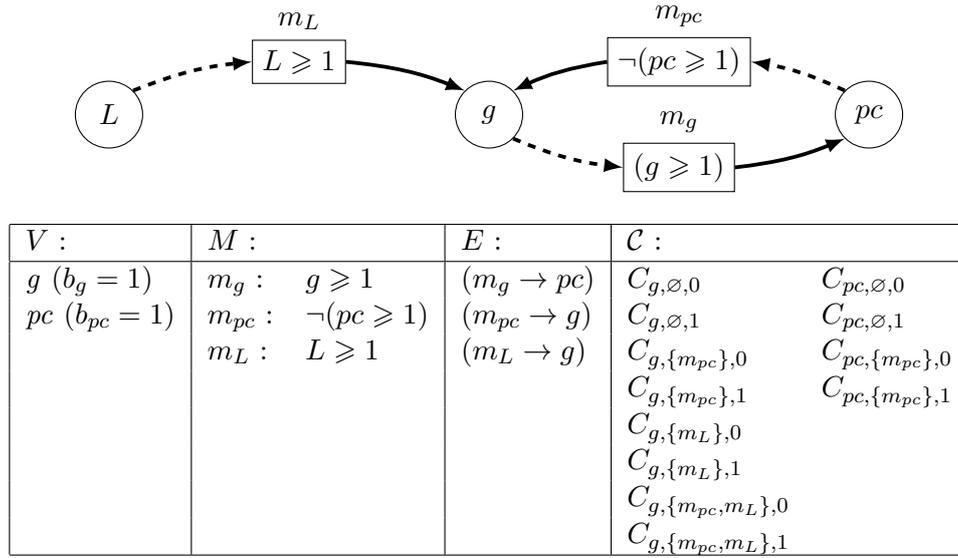


FIGURE 8.1 – Modèle simplifié de l’horloge circadienne présenté dans le chapitre 6.

dront, une fois créées, respectivement la liste des variables et la liste des multiplexes. Les deux parties suivantes permettent l’ajout d’une variable ou d’un zeitgeber, ainsi que leur niveau d’expression qualitatif maximal. La partie « CELE » contiendra la liste de toutes les célérités nécessaires à la simulation. Ces célérités seront initialisées à 0.

La dernière partie contient les boutons d’importation et d’exportation : « Display text only » génère le fichier de description du réseau au format SMBioNet [44], « Generate NetLogo Simulation » génère le fichier NetLogo de simulation, ainsi que l’interface graphique, et enfin, « Save model » sauvegarde le fichier de description du réseau au format XML, qui pourra être réimporté par la suite, à partir de la section « Upload model ».

Lors de l’ajout d’un multiplexe, une nouvelle partie apparaît et accompagne l’utilisateur pour construire la formule associée à ce multiplexe. La figure 8.3 présente les étapes de construction d’un multiplexe :

1. choisir le nom du multiplexe,
2. choisir le nom de la variable du premier atome,
3. choisir le comparateur,
4. choisir la valeur avec laquelle la variable sera comparée,
5. éventuellement cocher la case pour indiquer la négation de l’atome précédemment décrit,
6. répéter ces étapes pour tous les atomes dans le multiplexe,
7. choisir la ou les cibles de ce multiplexe.

Au fur et à mesure de la construction des multiplexes, le programme se charge de générer les paramètres et de les initialiser à 0. Le nom de chaque célérité est construit de la manière suivante :

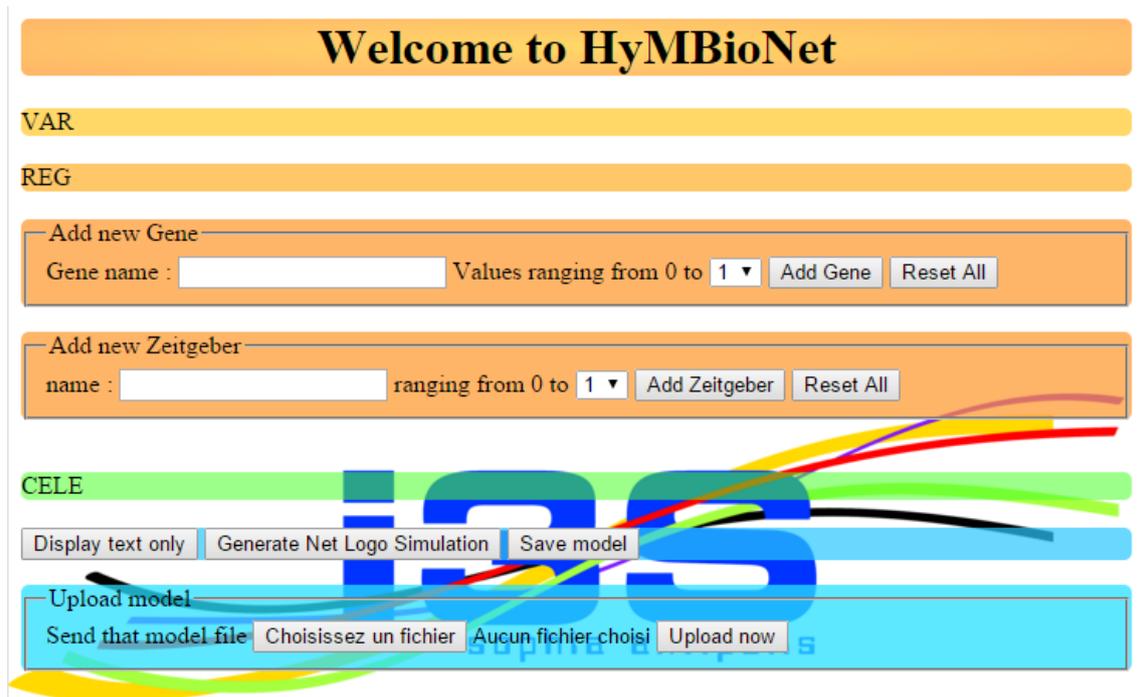


FIGURE 8.2 – Interface de HyMBioNet pour la description d’un réseau hybride de régulation génétique.

- le premier caractère est **C** pour « célérité »,
- la variable concernée est séparée du premier caractère par un underscore (-),
- la liste des ressources est contenue entre les crochets,
- le niveau de la variable est séparé de la liste des ressources par un underscore (-).

Par exemple, si nous avons une variable booléenne v qui a 2 prédécesseurs m_1 et m_2 , nous aurons 8 paramètres, $C_v[]_0$, $C_v[]_1$, $C_v[m1]_0$, $C_v[m1]_1$, $C_v[m2]_0$, $C_v[m2]_1$, $C_v[m1,m2]_0$ et $C_v[m1,m2]_1$. De manière générale, pour une variable v , il y a $(2^k \times (b_v + 1))$ paramètres où k est le nombre de prédécesseurs de la variable v .

Utilisons maintenant HyMBioNet sur notre exemple simplifié du cycle circadien (voir Section 7.3). En cliquant sur le bouton « Display text only » nous obtenons la figure 8.4. On y voit l’ensemble des variables « VAR » (divisé en zeitgebers et variables), l’ensemble des multiplexes « REG » et l’ensemble des célérités « CELE » générés automatiquement. La syntaxe de cette vue est inspirée de celle des fichiers d’entrée de SMBioNet et adaptée au formalisme hybride.

8.1.2 Interface de simulation

En utilisant le bouton « Generate NetLogo simulation », le programme génère un fichier `*.nlogo` contenant toutes les informations du réseau, ainsi que l’interface NetLogo. Pour pouvoir effectuer la simulation, il suffit de télécharger le logiciel NetLogo et de charger

Add new rule

Rule name :

Ml [

or

] => pc g

Adding L as left part of rule

Ml [L or

New Rule

] => pc g

Added : operator >=

Ml [L >=] => pc g

Adding 1 as right part of rule

Ml [Not ? Not or not Not ? (L >= 1)] => pc g

Checking for not : checked

Ml [!(L >= 1)] => pc g

Ml [!(L >= 1)] => pc

FIGURE 8.3 – Interface HyMBioNet de construction d'un multiplexe avec les différentes étapes.

Welcome to HyMBioNet

```

VAR
// Zeitgeber
L = 0 1;

// Genes
g = 0 1;
pc = 0 1;

REG
mg [g>=1] => pc
mpc [!(pc>=1)] => g
mL [L>=1] => g

CELE
C_g [_0 = 0
C_g [_1 = 0
C_g [mL]_0 = 0
C_g [mL]_1 = 0
C_g [mpc]_0 = 0
C_g [mpc]_1 = 0
C_g [mL,mpc]_0 = 0
C_g [mL,mpc]_1 = 0
C_pc [_0 = 0
C_pc [_1 = 0
C_pc [mg]_0 = 0
C_pc [mg]_1 = 0
back

```

ias
sophia antipalis

FIGURE 8.4 – Version textuelle du réseau de l'horloge circadienne de la figure 8.1.

le fichier dans le logiciel.

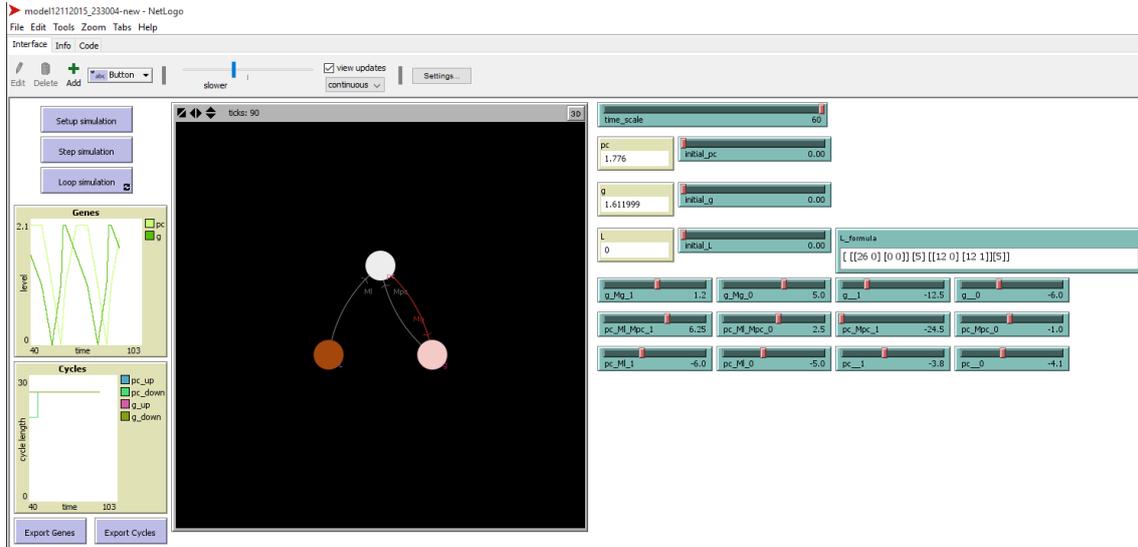


FIGURE 8.5 – Interface NetLogo de simulation du réseau simplifié du cycle circadien

La figure 8.5 présente l'interface NetLogo générée par HyMBioNet sur le réseau simplifié du cycle circadien. En haut à gauche se situent les boutons de lancement des simulations : « Setup simulation » permet d'initialiser la simulation, « Step simulation » fait avancer la simulation d'un pas de temps, et « Loop simulation » fait boucler la simulation indéfiniment. La grande fenêtre noire contient le graphe d'interaction du réseau dans lequel les couleurs des variables varient en fonction de leurs niveaux.

Le slider en haut à droite permet de changer le pas de temps entre 1 heure et 1 minute. Juste en dessous, nous avons un indicateur du niveau courant de chacune des variables, et un slider permettant de changer sa valeur initiale. Pour chacune des variables, la valeur du slider correspond à la somme de l'état discret et de la partie fractionnaire. Par exemple, la valeur 1 correspond à l'état discret 0 et la partie fractionnaire 1,0. Cette notation permet de masquer les doubles murs qui n'ont pas d'intérêt pour les simulations. En effet, la valeur 1 correspond aussi à $\eta(v) = 1$ et $\pi(v) = 0$.

La zone de texte à droite contient une description du scénario du zeitgeber pour la simulation. S'il y avait eu plusieurs zeitgebers, il y aurait eu autant de zones de texte. C'est ici que nous allons pouvoir spécifier le comportement des zeitgebers et obtenir, par exemple, une condition sans lumière, ou encore un décalage horaire. Le langage pour décrire le comportement des zeitgebers est défini comme suit :

- *formula* ::= [liste_cycles]
- *liste_cycles* ::= cycle | cycle liste_cycles
- *cycle* ::= [description][nb_cycles]
- *description* ::= phase | phase description
- *phase* ::= [heures valeur]
- *nb_cycles* est le nombre de cycles spécifiés par la *description* qui précède,

- *valeur* est la valeur discrète du zeitgeber,
- *heures* est le nombre d’heures que va passer le zeitgeber à la *valeur* donnée.

Par exemple, si nous voulons avoir une condition d’équinoxe, la formule décrivant le comportement du zeitgeber sera $[[[12\ 1]\ [12\ 0]]\ [3]]$: 12 heures à la valeur 1, puis 12 heures à la valeur 0. Cette alternance est répétée 3 fois. De la même façon, si nous voulons 4 cycles d’obscurité, puis 4 cycles d’équinoxe, la formule décrivant le comportement du zeitgeber sera $[[[24\ 0]]\ [4]\ [[12\ 1]\ [12\ 0]]\ [4]]$: 24 heures à 0 pendant 4 cycles, puis 12 heures à 1 et 12 heures à 0 pendant 4 cycles. Ce comportement du zeitgeber est répété indéfiniment.

La partie en bas à droite de la figure 8.5 contient les sliders permettant de régler les délais associés aux célérités. Chaque valeur est un réel représentant ce délai, accompagné d’un signe indiquant le sens de la célérité correspondante. Un délai de « -5 » donnera une célérité de $-\frac{1}{5}$.

L’interface contient également des graphiques, en bas à gauche, permettant d’afficher l’évolution du système. Le graphique du haut représente l’évolution des valeurs des variables au cours du temps, et celui du bas, l’évolution de la période. Les deux boutons « Export genes » et « Export cycles » permettent d’exporter ces données de simulation au format CSV.

8.1.3 Résultats pour le modèle de l’horloge

Dans cette section, nous comparons les simulations obtenues dans le chapitre précédent avec celles de HyMBioNet. Les figures suivantes (figure 8.6) sont produites par Gnuplot à partir des fichiers de simulation au format CSV. Les courbes bleues représentent les évolutions de la variable g , et les courbes rouges, celles de pc . Les zones grises correspondent aux phases d’obscurité. L’abscisse représente le temps et l’ordonnée la position dans l’espace des phases.

La figure 8.6 contient les simulations en condition d’obscurité constante (gauche) et d’équinoxe (droite). Nous pouvons y retrouver les deux glissements sur les murs extérieurs supérieurs de g ($\eta(g) = 1$ et $\pi(g) = 1, 0$) et de pc ($\eta(g) = 1$ et $\pi(g) = 1, 0$). Dans la condition d’obscurité constante, la batyphase de pc atteint le mur extérieur inférieur, contrairement à la condition d’équinoxe. Nous retrouvons ici la même différence entre ces deux conditions que dans le chapitre précédent.

La figure 8.7 contient les simulations montrant la résistance du modèle aux saisons. Dans la figure de gauche, la durée de la phase lumineuse est diminuée (16h de nuit et 8h de jour), alors que dans la figure de droite, elle est allongée (8h de nuit et 16h de jour). Les trajectoires sont très similaires dans ces deux conditions. Notons cependant que les états initiaux utilisés dans les deux conditions diffèrent : les initialisations correspondent à l’état

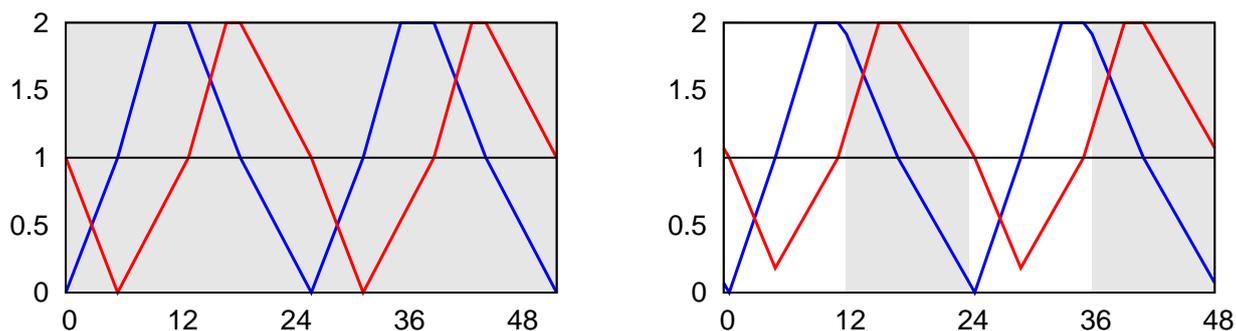


FIGURE 8.6 – Evolution des variables g (bleu) et pc (rouge) au cours du temps dans les conditions d'obscurité constante (droite) et d'équinox (gauche).

où se trouve le système lorsque le jour commence, une fois que la phase s'est stabilisée (voir Section 7.3.2 page 126). Puisque dans ces deux conditions, le rephasage automatique ne mène pas à la même phase stable, les initialisations pour ces simulations différent.

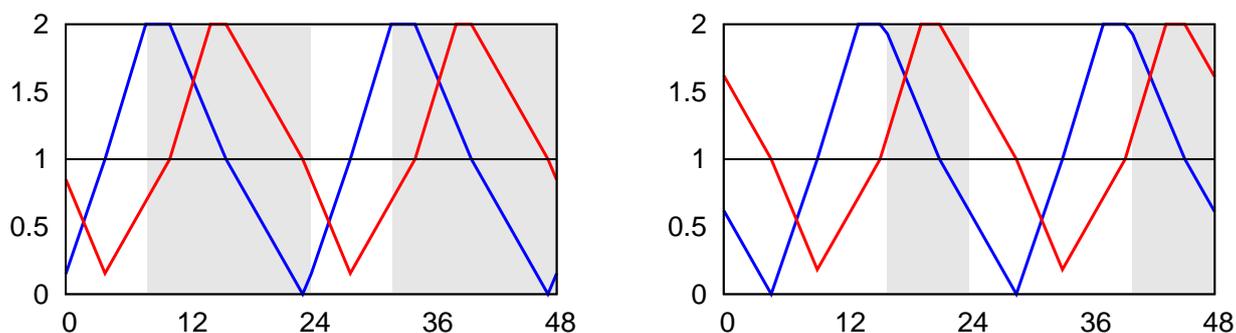


FIGURE 8.7 – Test de résistance aux saisons : évolution des variables g (bleu) et pc (rouge) au cours du temps dans les conditions de saison hivernale avec 16 heures d'obscurité et 8 heures de lumière (droite) et de saison estivale 8 heures d'obscurité et 16 heures de lumière (gauche).

Afin de tester le modèle dans les conditions du chapitre précédent (décalages horaires, durées extrêmes de phases de jour et de nuit), nous avons utilisé HyMBioNet et avons retrouvé les mêmes comportements. Cependant, la plus grande difficulté dans l'utilisation de HyMBioNet réside dans la détermination des nombreux paramètres.

8.2 Identification des contraintes sur les célérités

Dans le cadre hybride, l'identification des contraintes sur les célérités est d'autant plus compliquée qu'il existe une infinité de valeurs pour chacun des paramètres, contrairement au cadre de modélisation de R. Thomas. Dans le chapitre précédent, nous avons déjà déterminé certaines contraintes simples dans le cas d'un cycle limite en 2 dimensions.

Étant donné la taille des contraintes pour un modèle aussi simple, on s'attend à ce que les contraintes soient vraiment très grandes. Il est donc nécessaire d'avoir un moyen d'automatiser l'identification de ces contraintes. Une méthode formelle d'aide à l'identification des paramètres a été développée dans l'équipe par Jonathan Behaegel *et al.* [11, 9] : elle repose sur la notion de traces et sur la logique de Hoare.

8.2.1 Introduction à la logique de Hoare pour les réseaux génétiques

La logique de Hoare [40] a été introduite par Tony Hoare en 1969 pour donner un cadre formel dédié à la preuve de correction de programmes impératifs. Des règles logiques sont utilisées afin de prouver que les instructions s'enchaînent correctement en fonction de l'objectif du programme. Ainsi, chaque programme est représenté par un « triplet de Hoare » qui contient une précondition (états des variables avant le programme), un programme (suite d'instructions) et une postcondition (états des variables après le programme). Un triplet se représente donc par :

$$\{\text{Précondition}\} \text{ programme } \{\text{Postcondition}\}$$

Dans le cadre de l'aide à la modélisation des réseaux de régulations génétiques, cette logique a été détournée afin d'aider à la construction de contraintes sur les paramètres des modèles [16]. Nous donnerons ici essentiellement les intuitions de cette logique, sans entrer dans une description formelle.

Dans le cadre des modèles discrets de Thomas, les événements tels que les passages de seuils d'une variable peuvent être vus comme des instructions. Ainsi, un chemin dans la dynamique du réseau peut être interprété comme une suite d'instructions, c'est-à-dire un programme. La précondition et postcondition du triplet sont alors interprétées comme des conditions sur des états du système avant et après le chemin. En reprenant les idées de Dijkstra [27], le calcul de la plus faible précondition permet de déterminer les contraintes sur des paramètres discrets, de telle manière que le chemin spécifié devienne possible [14, 17].

Prenons le réseau du cycle circadien purement discret en condition d'obscurité constante (la variable L n'a aucun rôle). La recherche des paramétrisations menant à une oscillation en obscurité constante peut être facilitée par la logique de Hoare. En effet, nous pouvons spécifier un chemin représentant l'oscillation par le triplet suivant :

$$\left\{ \begin{array}{l} g = 0 \\ pc = 0 \end{array} \right\} \underbrace{g+}_{(4)} ; \underbrace{pc+}_{(3)} ; \underbrace{g-}_{(2)} ; \underbrace{pc-}_{(1)} \left\{ \begin{array}{l} g = 0 \\ pc = 0 \end{array} \right\}$$

La précondition et la postcondition sont les mêmes puisque, au bout d'un cycle, le système doit se retrouver au même état. Le chemin décrit les étapes pour obtenir un cycle : g

augmente en premier, puis pc augmente à son tour et enfin, les deux variables diminuent dans le même ordre.

Les contraintes sur les paramètres sont données par le calcul de la plus faible précondition. Il s'agit des conditions minimales nécessaires pour rendre possible le chemin du triplet à partir d'un état satisfaisant la postcondition. La méthode d'obtention de cette plus faible précondition consiste à déterminer successivement les conditions nécessaires, à chacune des étapes du chemin, en partant de la postcondition. Dans notre exemple :

- (1) comme la dernière étape est la diminution de pc , l'état précédent cette diminution est ($g = 0, pc = 1$). La contrainte sur le paramètre de pc s'appliquant à cet état s'écrit donc $K_{pc, \emptyset} = 0$. Si ce n'était pas le cas, pc ne pourrait pas diminuer.
- (2) L'avant-dernière étape est la diminution de g . L'état précédent est donc ($g = 1, pc = 1$). La contrainte liée à cette étape est $K_{g, \emptyset} = 0$.
- (3) L'étape d'avant est l'augmentation de pc , donc l'état précédent est ($g = 1, pc = 0$). La contrainte associée est $K_{pc, \{m_g\}} = 1$.
- (4) La première étape est l'augmentation de g , donc l'état précédent est ($g = 0, pc = 0$). La contrainte associée est $K_{g, \{m_{pc}\}} = 1$.

La conjonction de ces 4 contraintes permet de fixer tous les paramètres du réseau, tout en assurant l'existence du cycle spécifié. Cependant cette méthode d'identification n'est utilisable telle quelle que dans le cadre purement discret. Elle doit être étendue au cadre de modélisation hybride.

8.2.2 Extension de la logique de Hoare au cadre de modélisation hybride

L'extension de la logique de Hoare au cadre de modélisation hybride nécessite de décrire les chemins avec des informations temporelles [9]. Chaque instruction élémentaire correspond au passage d'un seuil, mais ce passage doit arriver après un certain délai, supposé issu d'expérimentations. De plus, les préconditions et postconditions doivent exprimer des contraintes, non seulement sur les états discrets, mais également sur les parties fractionnaires. Un triplet se présente donc comme suit :

$$\underbrace{\left\{ \begin{array}{l} D_{pre} \\ H_{pre} \end{array} \right\}}_{\text{précondition}} \underbrace{\left(\begin{array}{l} \delta_1 \\ assert_1 \\ v+ \end{array} \right); \left(\begin{array}{l} \delta_2 \\ assert_2 \\ u+ \end{array} \right)}_{\text{chemin}} \underbrace{\left\{ \begin{array}{l} D_{post} \\ H_{post} \end{array} \right\}}_{\text{postcondition}}$$

Les préconditions et postconditions se décomposent en conditions discrètes D_{\dots} (qui ne portent que sur les états discrets) et en contraintes hybrides H_{\dots} (qui font intervenir les parties fractionnaires). Chacune des instructions élémentaires contient les trois éléments suivants :

- un délai de traversé de l'état discret courant (δ). La somme de tous les délais du chemin correspondra au temps que mettra le modèle à parcourir tout le chemin,

- une assertion qui permet d'indiquer entre autres la présence ou l'absence de glissements le long d'un mur,
- la transition discrète qui s'effectuera après la traversée de l'état discret courant.

Dans cette approche, chaque instruction élémentaire correspond à l'enchaînement d'une transition continue et d'une transition discrète. Ainsi, chacune des instructions élémentaires permet de spécifier la traversée d'un état qualitatif à partir de l'état hybride d'entrée dans cet état jusqu'au passage du seuil suivant. C'est le choix qui a été fait pour assurer une uniformité entre les étapes élémentaires.

8.3 Identification des contraintes sur le modèle de l'horloge

Cette section est dédiée à l'utilisation de la logique de Hoare pour déterminer les contraintes du réseau génétique de l'horloge circadienne présenté dans le chapitre 6. La figure 8.8 contient la définition complète du réseau de régulation.

Afin de pouvoir appliquer la logique de Hoare sur notre exemple de l'horloge, il est nécessaire de considérer le zeitgeber comme une des variables du réseau. Comme le zeitgeber doit osciller, nous introduisons une nouvelle variable X , car la variable L prise seule ne peut pas générer d'oscillation, une boucle négative non triviale¹ étant nécessaire. Pour imposer un cycle au zeitgeber, il suffira de donner les délais correspondants. Par exemple, pour une oscillation d'équinoxe, il faudra donner une célérité de $\frac{1}{6}$ à chacune des célérités de la boucle négative entre X et L pour obtenir un cycle de 12 heures de jour et 12 heures de nuit :

- 6 heures d'augmentation de L , en posant $C_{L,\{m_X\},0} = \frac{1}{6}$;
- 6 heures d'augmentation de X , en posant $C_{X,\{m_{LX}\},0} = \frac{1}{6}$;
- 6 heures de diminution de L , en posant $C_{L,\emptyset,1} = \frac{1}{6}$;
- 6 heures de diminution de X , en posant $C_{X,\emptyset,1} = \frac{1}{6}$.

Ici, nous cherchons, dans le cas de la condition d'équinoxe, à identifier des contraintes sur les célérités *via* la logique de Hoare. Nous vérifions ensuite que les célérités choisies dans le chapitre précédent (Section 7.3) satisfont ces contraintes.

Tout d'abord, pour transcrire une trajectoire cyclique dans le triplet de Hoare, il est logique que la précondition et la postcondition soient identiques. Ces conditions correspondent à un point de référence du cycle et nous avons toute liberté pour le choisir. Nous faisons le choix de prendre l'état hybride correspondant à l'état sur lequel le passage de la nuit au jour se stabilise dans les simulations présentées dans le chapitre 6 : $\eta(g) = 0$, $\eta(pc) = 1$ et $\eta(L) = 1$ pour la condition discrète D , et $\pi(g) = 0,07$, $\pi(pc) = 0,07$ et $\pi(L) = 0$ pour la condition hybride H (point rouge dans la figure 8.9).

Il reste à écrire le chemin du triplet de Hoare. À partir de l'unique état satisfaisant

1. Une autorégulation négative donnerait un mur noir donc aucun cycle ne serait possible.

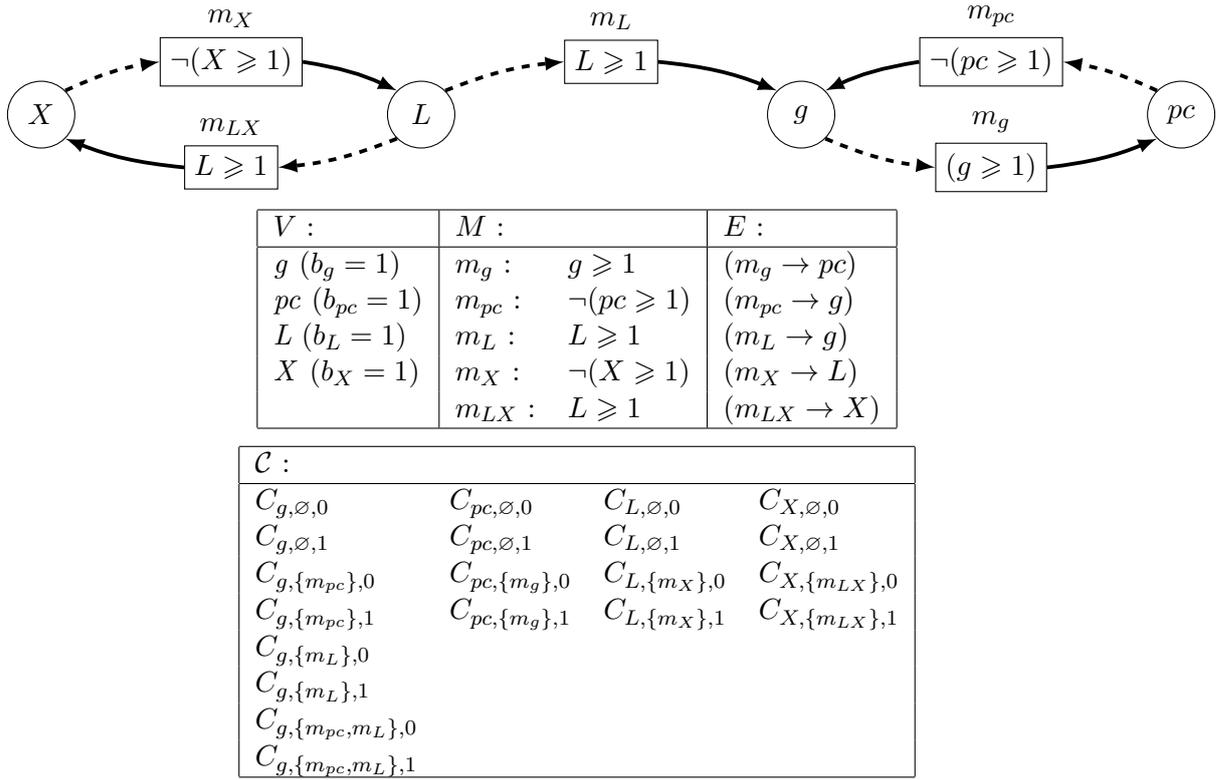


FIGURE 8.8 – Modèle simple de l’horloge circadienne présenté dans le chapitre 6 contenant également le zeitgeber L et la variable abstraite X nécessaire à son oscillation. Haut : graphe d’interaction du modèle. Bas : le quadruplet (V, M, E, \mathcal{C}) de ce réseau.

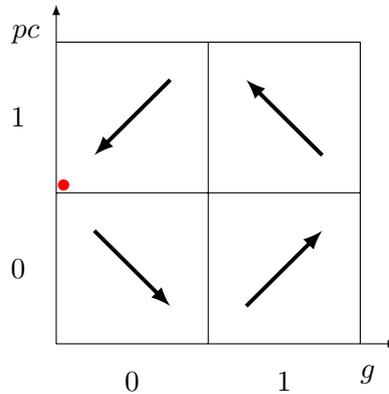


FIGURE 8.9 – Projection de l’espace des phases sur 2 dimensions. Le point rouge correspond à l’état hybride initial pour le triplet de Hoare.

la précondition, il faut décrire successivement les différentes instructions élémentaires du chemin. La première variable à passer un seuil est pc , elle va diminuer, la transition discrète est donc $pc-$. Comme nous ne voulons pas forcer un glissement à cet endroit du chemin, l’assertion sera l’atome \top . Finalement, nous imposons que le délai de l’état initial au passage du seuil de pc soit de 0,3 heure. Par conséquent, la première instruction élémentaire

du chemin est $\begin{pmatrix} 0,3 \\ \top \\ pc- \end{pmatrix}$. De la même façon, le prochain changement discret est l'augmentation de g , nous ne souhaitons pas non plus forcer de glissement et le délai voulu est 4,5. La deuxième instruction élémentaire est donc $\begin{pmatrix} 4,5 \\ \top \\ g+ \end{pmatrix}$. La construction de la suite du chemin est faite de la même façon. La notation $\text{slide}^+(g)$ et $\text{slide}^+(pc)$ indique des glissements imposés sur les murs extérieurs de g et de pc .

Remarque : un glissement a été imposé sur le mur extérieur inférieur de la variable X (noté $\text{slide}^-(X)$) pour avoir un unique cycle limite pour l'oscillation de la lumière (voir la propriété démontrée dans la section 6.2.2 du chapitre 6).

Après avoir déterminé toutes les instructions élémentaires de notre chemin, nous obtenons le triplet suivant :

$$\left\{ \begin{matrix} D \\ H \end{matrix} \right\} \begin{pmatrix} 0,3 \\ \top \\ pc- \end{pmatrix} ; \begin{pmatrix} 4,5 \\ \top \\ g+ \end{pmatrix} ; \begin{pmatrix} 1,2 \\ \top \\ X+ \end{pmatrix} ; \begin{pmatrix} 5,1 \\ \text{slide}^+(g) \\ pc+ \end{pmatrix} ; \begin{pmatrix} 0,9 \\ \top \\ L- \end{pmatrix} ; \begin{pmatrix} 5,2 \\ \text{slide}^+(pc) \\ g- \end{pmatrix} ; \begin{pmatrix} 0,8 \\ \text{slide}^-(L) \\ X- \end{pmatrix} ; \begin{pmatrix} 6 \\ \top \\ L+ \end{pmatrix} \left\{ \begin{matrix} D \\ H \end{matrix} \right\}$$

Nous pouvons remarquer que la somme des délais est égale à 24 heures. Le glissement imposé sur L a pour unique fonction d'obtenir une oscillation entretenue de la lumière (comme expliqué dans le chapitre 6).

Le calcul de la plus faible précondition a été implémenté. Le programme prend en entrée un fichier contenant le réseau et le triplet de Hoare, et renvoie des contraintes sur les célérités du réseau. Les contraintes obtenues sont présentées dans la table 8.1, qui contient aussi les valeurs choisies pour le modèle du chapitre précédent. Le domaine des célérités étant continu, il existe une infinité de combinaisons possibles de célérités permettant au modèle d'exhiber le comportement que nous souhaitons. Les célérités du chapitre précédent respectant toutes les contraintes, elles constituent une de ces combinaisons.

L'utilisation de la logique de Hoare nous a donc permis de montrer formellement que le modèle admet une trajectoire cyclique dans des conditions d'équinoxe. La section suivante va encore plus loin en construisant les contraintes sur les 56 célérités d'un modèle réaliste du cycle cellulaire.

Célérités	Contraintes	valeurs choisies au chapitre précédent
$C_{g,\emptyset,0}$	< 0	$-\frac{1}{7.5} = -0.13$
$C_{g,\emptyset,1}$	$\in [-0.19, 0]$	$-\frac{1}{5.5} = -0.18$
$C_{g,\{m_{pc}\},0}$	pas de contrainte	$\frac{1}{5.5} = 0.18$
$C_{g,\{m_{pc}\},1}$	pas de contrainte	$\frac{1}{4} = 0.25$
$C_{g,\{m_L\},0}$	$\in [-3.32, 0]$	$-\frac{1}{7.5} = -0.13$
$C_{g,\{m_L\},1}$	$\in [0, 1.11]$	$-\frac{1}{10} = -0.1$
$C_{g,\{m_{pc},m_L\},0}$	$\in [0, 0.22]$	$\frac{1}{4.5} = 0.22$
$C_{g,\{m_{pc},m_L\},1}$	> 0	$\frac{1}{4} = 0.25$
$C_{pc,\emptyset,0}$	< 0	$-\frac{1}{5.5} = -0.18$
$C_{pc,\emptyset,1}$	$\in [-0.14, 0]$	$-\frac{1}{7.5} = -0.13$
$C_{pc,\{m_g\},0}$	$\in [0, 0.16]$	$\frac{1}{7.5} = 0.13$
$C_{pc,\{m_g\},1}$	> 0	$\frac{1}{4} = 0.25$

TABLE 8.1 – Contraintes obtenues pour chacune des célérités et valeurs choisies au chapitre précédent.

8.4 Identification des contraintes pour un modèle du cycle cellulaire

Le cycle cellulaire est un des mécanismes biologiques les plus importants du vivant, puisqu'il contrôle la multiplication des cellules. Un dysfonctionnement de ce cycle peut mener à un cancer dans lequel la prolifération cellulaire devient anormale. Il est ainsi largement étudié : un grand nombre de modèles ont été construits en utilisant différents formalismes, que ce soit des formalismes continus [55, 32, 24, 34], hybrides [63] ou discrets [59, 60]. Cette section présente un nouveau modèle hybride du cycle cellulaire présenté dans [10], dont les paramètres ont été déterminés grâce au calcul de la plus faible précondition d'un triplet de Hoare présenté dans la section précédente.

8.4.1 Le réseau du cycle cellulaire

Le cycle cellulaire est contrôlé par un réseau de régulation génétique qui permet à la cellule de se diviser en deux cellules filles identiques. Sa dynamique se décompose en 4 phases :

Phase G1 Les facteurs de croissance présents dans la cellule lui permettent de prendre du volume. La cellule doit atteindre une taille suffisante pour que, après la division, les deux cellules filles aient une taille suffisante pour être viables. Il s'agit de la phase

la plus longue : environ 10 heures. Du point de vue moléculaire, cette phase est caractérisée par la présence du complexe CyclineD-CDK4/6 qui est activé par les facteurs de croissance. La phase G1 se termine par un point de contrôle permettant le passage de la phase G1 à la phase S.

Phase S Cette phase est déclenchée par le complexe CyclineE-CDK2 qui permet le passage du point de contrôle G1/S. La phase S correspond à la réplication du génome de la cellule : elle passe de $2n$ chromosomes à $4n$, de façon à ce que les deux cellules filles aient chacune $2n$ chromosomes. Au niveau moléculaire, CyclineE-CDK2 active le complexe CyclineA-CDK2.

Phase G2 La cellule termine la réplication de l'ADN et se prépare à la mitose (division cellulaire). Le complexe CyclineA-CDK2 active le complexe CyclineB-CDK1 et, si l'ADN est bien répliqué et n'a subi aucun dommage, le cycle passe le point de contrôle G2/M qui engage la mitose.

Phase M La cellule mère se divise en deux cellules filles identiques génétiquement.

Certains complexes impliqués dans le cycle cellulaire sont exprimés en même temps dans la cellule. Ils sont alors régulés par des kinases pour être actifs ou inactifs pendant certaines phases (les kinases CDC25 ou CAK pour l'activation, et p21 ou p27 pour l'inhibition).

Le modèle du cycle cellulaire construit dans l'article est inspiré du modèle de John Tyson [75] et contient 5 variables :

- **SK** correspond à des complexes qui contrôlent le passage de G1 à S. En particulier, l'autorégulation de SK, lorsqu'elle est active correspond à ce point de contrôle.
- **A** correspond à la cyclineA. Une valeur élevée de cette variable signifie que le cycle est en phase S.
- **B** représente la cyclineB. Une valeur élevée de cette variable signifie que le cycle est en phase G2.
- **EP** représente le point de contrôle entre les phases G2 et M et contribue à l'inhibition de B pour terminer la phase G2.
- **En** représente les inhibiteurs p21 et p27 qui permettent de mettre fin aux phases G2 et S.

La figure 8.10 contient le graphe d'interaction de ce réseau. Avec ses 5 variables, ce réseau a 56 célérités différentes (voir Table 8.2). Le calcul de la plus faible précondition sur un triplet de Hoare va donc nous être très utile pour déterminer des paramètres réalistes.

8.4.2 Triplet de Hoare

Contrairement à l'exemple précédent, le chemin discret emprunté par le cycle cellulaire (l'ordre des changements qualitatifs) n'est pas connu de manière non ambiguë. Cela nous oblige à considérer tous les chemins possibles compatibles avec les observations. Celui présenté dans l'article (et ci-dessous) est un chemin parmi les 12 possibles.

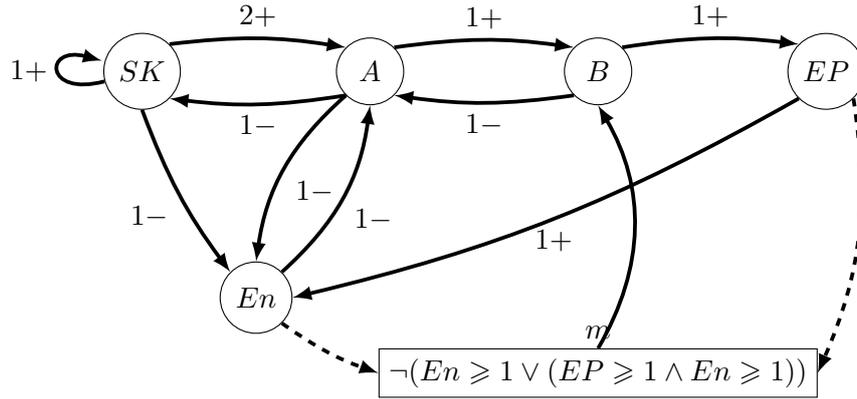


FIGURE 8.10 – Graphe d'interaction du modèle du cycle cellulaire à 5 variables. La notation $v \xrightarrow{s+} u$ représente le multiplexe $\boxed{v \geq s} \rightarrow u$ dont le nom est aussi v par abus de notation.

SK	A	B	EP	En
$C_{SK,\emptyset,0}$	$C_{A,\emptyset,0}$	$C_{B,\emptyset,0}$	$C_{EP,\emptyset,0}$	$C_{En,\emptyset,0}$
$C_{SK,\emptyset,1}$	$C_{A,\emptyset,1}$	$C_{B,\emptyset,1}$	$C_{EP,\emptyset,1}$	$C_{En,\emptyset,1}$
$C_{SK,\emptyset,2}$	$C_{A,\{SK\},0}$	$C_{B,\{A\},0}$	$C_{EP,\{B\},0}$	$C_{En,\{SK\},0}$
$C_{SK,\{A\},0}$	$C_{A,\{SK\},1}$	$C_{B,\{A\},1}$	$C_{EP,\{B\},1}$	$C_{En,\{SK\},1}$
$C_{SK,\{A\},1}$	$C_{A,\{B\},0}$	$C_{B,\{m\},0}$		$C_{En,\{A\},0}$
$C_{SK,\{A\},2}$	$C_{A,\{B\},1}$	$C_{B,\{m\},0}$		$C_{En,\{A\},1}$
$C_{SK,\{SK\},0}$	$C_{A,\{En\},0}$	$C_{B,\{A,m\},0}$		$C_{En,\{EP\},0}$
$C_{SK,\{SK\},1}$	$C_{A,\{En\},1}$	$C_{B,\{A,m\},0}$		$C_{En,\{EP\},1}$
$C_{SK,\{SK\},2}$	$C_{A,\{SK,B\},0}$			$C_{En,\{SK,A\},0}$
$C_{SK,\{A,SK\},0}$	$C_{A,\{SK,B\},1}$			$C_{En,\{SK,A\},1}$
$C_{SK,\{A,SK\},1}$	$C_{A,\{SK,En\},0}$			$C_{En,\{SK,EP\},0}$
$C_{SK,\{A,SK\},2}$	$C_{A,\{SK,En\},1}$			$C_{En,\{SK,EP\},1}$
	$C_{A,\{B,En\},0}$			$C_{En,\{A,EP\},0}$
	$C_{A,\{B,En\},1}$			$C_{En,\{A,EP\},1}$
	$C_{A,\{SK,B,En\},0}$			$C_{En,\{SK,A,EP\},0}$
	$C_{A,\{SK,B,En\},1}$			$C_{En,\{SK,A,EP\},1}$

TABLE 8.2 – Célérités du réseau de la figure 8.10.

Le triplet de Hoare utilisé dans cette démarche est le suivant :

$$\left\{ \begin{array}{l} SK = 0 \\ EP = 0 \\ A = 0 \\ B = 0 \\ En = 1 \end{array} \right\} \underbrace{\left(\begin{array}{c} t_1 \\ S^-(EP) \\ SK+ \end{array} \right); \left(\begin{array}{c} t_2 \\ \top \\ SK+ \end{array} \right); \left(\begin{array}{c} t_3 \\ S^+(SK) \\ En- \end{array} \right); \left(\begin{array}{c} t_4 \\ S^-(En) \\ A+ \end{array} \right); \left(\begin{array}{c} t_5 \\ \top \\ SK- \end{array} \right); \left(\begin{array}{c} t_6 \\ S^+(A) \\ SK- \end{array} \right); \left(\begin{array}{c} t_7 \\ S^-(SK) \\ B+ \end{array} \right); \dots }_{\substack{G1 \\ t_1 + t_2 + t_3 = 10h}} \underbrace{\left(\begin{array}{c} t_4 \\ S^-(En) \\ A+ \end{array} \right); \left(\begin{array}{c} t_5 \\ \top \\ SK- \end{array} \right); \left(\begin{array}{c} t_6 \\ S^+(A) \\ SK- \end{array} \right); \left(\begin{array}{c} t_7 \\ S^-(SK) \\ B+ \end{array} \right); \dots }_{\substack{S \\ t_4 + t_5 + t_6 + t_7 = 8h}}$$

$$\dots \underbrace{\left(\begin{array}{c} t_8 \\ \top \\ A- \end{array} \right); \left(\begin{array}{c} t_9 \\ S^+(B) \\ EP+ \end{array} \right); \left(\begin{array}{c} t_{10} \\ S^+(EP) \\ En+ \end{array} \right)}_{\text{G2}} \underbrace{\left(\begin{array}{c} t_{11} \\ S^-(A) \wedge S^+(En) \\ B- \end{array} \right); \left(\begin{array}{c} t_{12} \\ S^-(B) \\ EP- \end{array} \right)}_{\text{M}} \left\{ \begin{array}{l} SK = 0 \\ EP = 0 \\ A = 0 \\ B = 0 \\ En = 1 \end{array} \right.$$

$$t_8 + t_9 = 4h \qquad t_{10} + t_{11} + t_{12} = 0.5h$$

Nous n'avons pas toutes les informations sur les délais de chacun des états discrets. En effet, la durée de chacune des phases du cycle cellulaire est connue, mais les délais associés à chacun des états discrets de chacune des phases sont inconnus. Par exemple, la phase G1 est représentée par 3 états discrets dont les délais sont notés t_1 , t_2 et t_3 . Nous savons seulement que la phase G1 dure 10 heures, donc que $t_1 + t_2 + t_3 = 10h$. Ici, la répartition de ces 10 heures a été faite de manière équitable sur les 3 états discrets. Nous considérons donc que $t_1 = t_2 = t_3 = 3,33h$. Les délais associés à chacun des états discrets pour les autres phases sont obtenus avec des hypothèses similaires. La table 8.3 donne les contraintes obtenues grâce au calcul de la plus faible précondition.

En	$ \begin{aligned} C_{En,\{ \},0} &< 0 \\ C_{En,\{ \},1} &< 0 \\ C_{En,\{A\},0} &< -\frac{4}{\Delta t_S} \\ C_{En,\{A\},1} &= -\frac{4}{2 * \Delta t_{G1}} \\ C_{En,\{SK\},0} &< 0 \\ C_{En,\{SK\},1} &< 0 \\ C_{En,\{A,SK\},0} &< \frac{2}{\Delta t_{G2}} \\ C_{En,\{A,SK\},1} &> 0 \\ C_{En,\{A,EP,SK\},0} &= \frac{6 - 3 * \Delta t_{G2} * C_{En,\{A,SK\},0}}{2 * \Delta t_M} \\ C_{En,\{A,EP,SK\},1} &> \frac{\Delta t_M}{3} \end{aligned} $	EP	$ \begin{aligned} C_{EP,\{ \},0} &< -\frac{3}{\Delta t_{G1}} \\ C_{EP,\{ \},1} &= -\frac{3}{\Delta t_M} \\ C_{EP,\{B\},0} &= \frac{1}{\Delta t_{G2}} \\ C_{EP,\{B\},1} &> \frac{3}{\Delta t_M} \end{aligned} $
		B	$ \begin{aligned} C_{B,\{ \},0} &< -\frac{3}{\Delta t_M} \\ C_{B,\{ \},1} &= -\frac{3}{\Delta t_M} \\ C_{B,\{m\},0} &< \frac{4}{\Delta t_S} \\ C_{B,\{m\},1} &> \frac{2}{\Delta t_{G2}} - C_{B,\{A,m\},1} \\ C_{B,\{A,m\},0} &= \frac{4 - \Delta t_S * C_{B,\{m\},0}}{3 * \Delta t_S} \\ C_{B,\{A,m\},1} &< \frac{2}{\Delta t_{G2}} \end{aligned} $
A	$ \begin{aligned} C_{A,\{ \},0} &< -\frac{6 - (3 * \Delta t_{G2} + 2 * \Delta t_M) * C_{A,\{En\},0}}{2 * \Delta t_M} \\ C_{A,\{ \},1} &< 0 \\ C_{A,\{B\},0} &< 0 \\ C_{A,\{B\},1} &< 0 \\ 0 &> C_{A,\{En\},0} > -\frac{2}{\Delta t_{G2}} \\ C_{A,\{En\},1} &= -\frac{2}{\Delta t_{G2}} \\ C_{A,\{B,En\},0} &> 0 \\ C_{A,\{B,En\},1} &> \frac{4}{\Delta t_S} - C_{A,\{B,En,SK\},1} > 0 \\ C_{A,\{B,SK\},0} &< 0 \\ C_{A,\{B,SK\},1} &< 0 \\ C_{A,\{B,En,SK\},0} &= \frac{4}{\Delta t_S} \\ 0 &< C_{A,\{B,En,SK\},1} < \frac{4}{\Delta t_S} \end{aligned} $	SK	$ \begin{aligned} C_{SK,\{ \},0} &< -\frac{4}{\Delta t_S} \\ C_{SK,\{A\},0} &= \frac{4}{2 * \Delta t_{G1} + 6 * \Delta t_M + 3 * \Delta t_{G2}} \\ C_{SK,\{SK\},1} &= -\frac{\Delta t_S}{4} \\ C_{SK,\{SK\},2} &= -\frac{4}{\Delta t_S} \\ C_{SK,\{A,SK\},1} &= \frac{3}{\Delta t_{G1}} \\ C_{SK,\{A,SK\},2} &> \frac{3}{\Delta t_{G1}} \end{aligned} $

TABLE 8.3 – Contraintes sur les célérités obtenues en utilisant le calcul de la plus faible précondition où Δt_{G1} (resp. Δt_S , Δt_{G2} , Δt_M) est la durée de la phase G1 (resp. S, G2, M).

8.4.3 Simulation

La simulation suivante a été obtenue en choisissant des valeurs de célérités satisfaisant les contraintes de la section précédente. Ce choix est arbitraire, et d'autres valeurs de célérités satisfaisant les mêmes contraintes donneraient des simulations semblables.

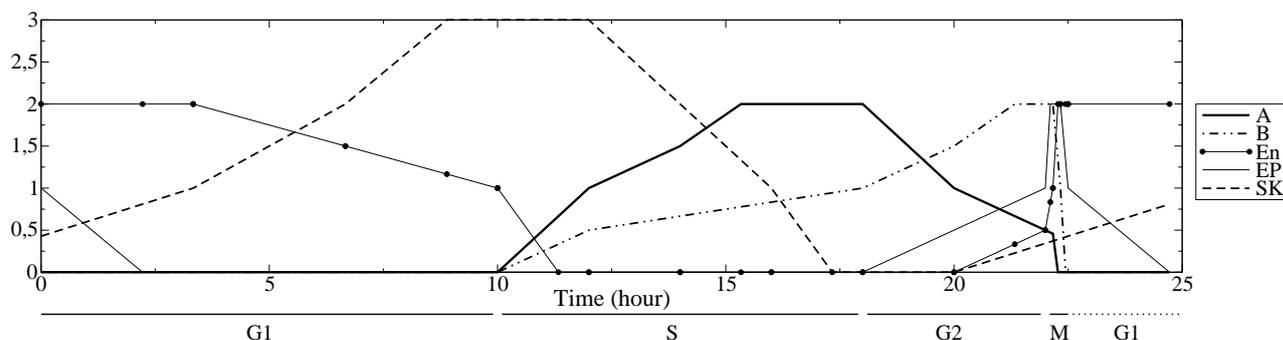


FIGURE 8.11 – Simulation du modèle du cycle cellulaire

8.5 Conclusion

Depuis que nous défini le cadre de modélisation hybride de réseaux de régulation génétique, un certain nombre de travaux poursuivent cette approche, d'une part pour faciliter les échanges avec les experts des domaines biologiques, et d'autre part, en vue d'une automatisation du processus de modélisation.

Tout d'abord, la plateforme *HyMBioNet* a été développée en collaboration avec Gilles Enée. Cette plateforme est très utile dans les discussions avec les biologistes, puisqu'elle rend possibles des simulations du modèle d'étude en direct. L'interface graphique permet de changer facilement les paramètres et d'observer l'effet qu'ont ces changements sur le comportement du modèle. L'inconvénient principal de l'utilisation du logiciel *NetLogo* réside dans la discrétisation du temps. En effet, *HyMBioNet* discrétise le temps en intervalles réguliers. À un instant donné, le calcul de la partie fractionnaire suivante en fonction de la célérité peut mener à la sortie de l'intervalle $[0,1]$, ce qui mène à des approximations limitantes. Pour éviter que l'accumulation de ces approximations entraîne des trajectoires qualitativement différentes des celles du modèle, *HyMBioNet* contrôle ces sorties de l'intervalle $[0,1]$ pour donner à l'utilisateur des simulations exploitables.

Une méthode formelle d'aide à l'identification des célérités a été également développée. Cette méthode est basée sur la logique de Hoare qui a été adaptée à la modélisation des réseaux génétiques hybrides. Le principe réside dans la détermination d'un ensemble de contraintes sur les célérités d'un réseau. Lorsque l'on choisit des célérités satisfaisant ces contraintes, le modèle exhibe la trace spécifiée représentant un comportement biologique.

La principale difficulté pour le modélisateur est de connaître un comportement biologique à traduire sous forme de triplet de Hoare : si le modélisateur ne connaît pas les détails du chemin, il peut avoir de grandes difficultés à écrire le triplet. Dans ces cas-là, il utilisera les traces issues d'expérimentations biologiques pour extraire les informations manquantes. Cette difficulté dépassée, un solveur de contraintes donnera des jeux de paramètres compatibles avec les connaissances biologiques.

Chapitre 9

Conclusion et perspectives

Un grand nombre de mécanismes biologiques sont contrôlés par des réseaux génétiques dans lesquels des produits de gènes interagissent avec d'autres gènes créant une dynamique complexe. Aborder la dynamique de ces mécanismes, souvent très imposants, est difficile sans outils formels de modélisation. Dans cette thèse, nous avons fait le choix d'utiliser le formalisme de René Thomas et de l'étendre, en raison de son adéquation remarquable avec les méthodes formelles de l'informatique.

9.1 Contributions

Le formalisme discret de René Thomas duquel nous sommes partis permet de représenter facilement la dynamique globale d'un réseau. Son faible nombre de valeurs possibles de paramètres et son aspect formel permettent d'utiliser des méthodes d'identification automatique des valeurs des paramètres, en se référant aux comportements biologiques connus. Cette approche a été mise en œuvre à partir de 2004 avec le logiciel SMBioNet qui, au travers de l'énumération de l'ensemble des combinaisons de valeurs des paramètres, permet leur identification automatique [15].

Dans cette lignée, nous avons proposé deux contributions, l'une qui exploite les méthodes formelles pour étendre les travaux d'Aurélien Naldi et faciliter la réduction des réseaux de Thomas tout en préservant les comportements asymptotiques, l'autre qui enrichit l'approche discrète de Thomas par la prise en compte d'un temps continu, et débouche sur une version hybride du formalisme. À l'occasion de cette seconde contribution, nous proposons un modèle du cycle circadien chez les mammifères fidèle à la plupart des propriétés connues tout en étant d'une simplicité remarquable.

9.1.1 Réduction formelle de réseaux de régulation génétique

Dans cette contribution, d'une part, nous avons apporté, grâce aux multiplexes, un cadre formel à l'approche de Naldi, et d'autre part, nous avons développé une extension permettant la suppression de seuils. Nous avons de ce fait formalisé les idées initiales de Naldi en considérant comme lui que la variable supprimée évolue vite. La suppression d'une variable revient alors à connecter ses prédécesseurs à ses successeurs de manière adéquate. Nous avons également dépassé les acquis offerts par la méthode de Naldi en permettant la suppression de seuils sur des variables autorégulées, pourvu que ces seuils ne soient pas ceux des autorégulations. L'idée reste similaire à celle de Naldi : la variable concernée est supposée traverser rapidement le seuil supprimé.

Nous avons ensuite démontré rigoureusement que toutes nos réductions successives assurent la conservation des comportements asymptotiques (états stables et attracteurs cycliques). Du point de vue biologique, ces comportements sont, d'une part, très riche

d'information, car ils mettent en évidence des états particuliers du système biologique (états pathologiques, épigénèses, homéostasies ...), et d'autre part, ils ne requièrent pas de lourds protocoles expérimentaux pour être observés.

Enfin, nous avons illustré notre nouveau formalisme de réduction, en établissant le lien entre deux modèles qui avaient été conçus indépendamment (l'un à 4 variables et l'autre à 2 variables). Il s'agit du réseau classique responsable du switch génétique du phage λ . Nous avons établi que le second est une abstraction valide, par réductions successives, du premier. Nous avons ensuite effectué des réductions supplémentaires qui ont mené à un modèle à 2 variables booléennes. Au bilan, on passe d'un modèle à 48 états, à un modèle à 4 états exhibant les mêmes comportements asymptotiques.

9.1.2 Cadre de modélisation hybride

Dans cette contribution, nous avons voulu préserver l'état d'esprit apporté par Thomas et Snoussi avec les paramètres discrets $K_{v,\omega}$. Afin de refléter dans un modèle de réseau génétique l'écoulement d'un temps continu, nous avons remplacé ces paramètres discrets par des paramètres réels que nous avons appelés des *célérités*. Ces célérités dépendent également de l'état de la variable afin de mieux représenter le fait que le temps de traversée d'un état discret dépend d'éléments additionnels aux ressources comme la distance au point focal.

Ces enrichissements de la théorie de Thomas et Snoussi offrent un formalisme hybride qui résout plusieurs problèmes rencontrés par les formalismes hybrides antérieurs qui nécessitaient la plupart du temps des remises à zéro d'horloges et échouaient, entre autres, à modéliser les phénomènes d'accumulation. De plus, de même que les paramètres K ont permis une représentation compacte de la dynamique d'un réseau de Thomas, nous avons éliminé, grâce aux célérités, bon nombre de redondances ce qui rend le formalisme plus élégant.

Ce formalisme a été motivé par la modélisation du cycle circadien et la puissance offerte par les célérités a abouti à un modèle mettant en jeu un nombre de variables remarquablement restreint, tout en conservant la crédibilité biologique de la dynamique. Dans ce cadre et en limitant le nombre de variables à 2, nous avons proposé une première méthode d'identification des célérités qui assure un cycle limite. La méthode utilisée démontre qu'au moins une des variables doit glisser sur un mur extérieur. Les contraintes obtenues par cette méthode ne permettent pas d'identifier toutes les célérités et la connaissance de comportements saisonniers ou à l'issue d'un décalage horaire nous ont permis d'identifier les célérités manquantes. Pour cela, il a été nécessaire d'effectuer des simulations en assez grand nombre et nous avons développé le logiciel `CircaPlot` à cet effet.

In fine, le modèle obtenu montre un comportement très réaliste malgré sa simplicité. En

effet, lors de perturbations telles que les variations saisonnières ou les décalages horaires, le modèle adapte sa phase et sa période en un nombre de jours compatible avec l'expérience commune. En particulier dans le cas des décalages horaires, la différence du temps de resynchronisation du système lors d'une avance de phase est qualitativement plus longue que lors d'un retard de phase.

La simplicité du modèle lui permet d'être utilisé comme un module dans des modèles plus gros dont le mécanisme biologique correspondant est dépendant du cycle circadien. Sa faible taille et le fait que de nombreux paramètres soient déjà fixés facilitent grandement son intégration et limite l'augmentation du nombre de paramètres à identifier dans le modèle global.

Plus généralement, au cours de cette thèse, d'autres travaux ont réutilisé notre formalisme :

- La plateforme de simulation `HyMBioNet` a été développée dans le but de faciliter les discussions avec les biologistes. Elle possède deux parties : une partie de construction du modèle dans laquelle une interface graphique assiste l'utilisateur dans la définition du réseau, et une partie de simulation proposant une interface graphique fondée sur `NetLogo`. Malgré `HyMBioNet`, le logiciel `CircaPlot` est resté indispensable pour étudier les décalages de phases.
- À la suite des premiers résultats obtenus en 2 dimensions pour le cycle circadien, une méthode d'identification des célérités a été développée. Cette méthode s'inspire de la logique de Hoare pour déterminer les contraintes permettant au modèle d'adopter un comportement temporel donné.

Ces divers travaux de recherche, qui réutilisent tous le formalisme hybride développé dans cette thèse, démontrent sa bonne adéquation pour aborder des problèmes biologiques dans lesquels le temps joue un rôle important. C'est à l'occasion du développement d'autres modèles de réseaux biologiques fondés sur notre théorie que des collègues ont effectué ces extensions.

9.2 Perspectives

9.2.1 Réduction symbolique

Le but de la réduction de réseaux de régulations génétiques est de permettre l'identification des paramètres cohérents avec la biologie même lorsque le réseau est d'une taille ne permettant pas l'énumération des paramètres et la méthode présentée dans cette thèse constitue le premier pas vers cet objectif. Avec notre méthode, pour réduire un réseau, il est nécessaire de connaître la valeur des paramètres or, généralement, les paramètres ne sont pas connus puisque nous cherchons à les identifier. Nous pouvons alors envisager une méthode dont tous les paramètres sont des symboles et non des entiers. La réduction symbolique d'un réseau consisterait alors à définir un autre réseau, réduit, dans lequel les

nouveaux paramètres s'exprimeraient en fonction des paramètres symboliques du réseau initial. La mémorisation des liens entre les différentes étapes de réduction pourrait alors permettre de retrouver le réseau initial à partir du réseau réduit.

Cette mémorisation permettrait alors de créer un lien formel entre les paramètres du réseau initial et ceux du réseau réduit. Si les paramètres du réseau réduit sont connus, il devient possible de construire des contraintes sur les paramètres du réseau initial pour que les comportements asymptotiques soient compatibles. Une stratégie de modélisation émerge alors :

1. On commence par effectuer une réduction symbolique du réseau initial afin d'obtenir le plus petit réseau possible.
2. Sur ce réseau réduit, on effectue une identification des modèles compatibles avec la biologie par model checking en ne retenant que les modèles qui sont en accord avec les connaissances biologiques.
3. On « remonte » les réductions effectuées afin d'extraire, pour chacun des modèles du réseau réduit, tous les modèles du réseau initial compatibles.

Cette méthode permettrait alors d'avoir un ensemble exhaustif des modèles compatibles avec les connaissances biologiques, sans avoir à énumérer l'intégralité des combinaisons de paramètres. Ainsi, nous pouvons limiter l'explosion combinatoire qui constitue la limite de l'identification des paramètres par le model checking.

Il serait également largement utile de généraliser la méthode de réduction afin de préserver des dynamiques autres que les comportements asymptotiques. Dans la méthode de réduction actuelle, nous priorisons des transitions afin de simplifier le modèle plus facilement. Cela supprime un certain nombre d'informations et l'atteignabilité n'est généralement pas conservée au cours du processus. Nous pourrions alors suivre une autre voie de simplification : spécifier les chemins que l'on souhaite conserver au cours des réductions et identifier les réductions compatibles avec ces comportements. Les chemins spécifiés pourraient prendre la forme d'un triplet de Hoare par exemple. Il faudrait alors coupler la réduction du réseau et une réécriture du triplet de Hoare pour garder une cohérence entre les deux.

9.2.2 Modélisation hybride

Dans le formalisme hybride, un certain nombre de détails sont encore à développer.

Par exemple, les contraintes identifiées dans le chapitre 6 ne sont applicables qu'en deux dimensions. Il faudrait donc généraliser à n'importe quelle dimension pour qu'elles puissent être utilisées dans des modèles plus grands.

De la même manière que Houssine Snoussi a montré la compatibilité du cadre de modélisation de Thomas avec les systèmes d'équations différentielles linéaires par mor-

ceaux, nous pourrions chercher à identifier les comportements des modèles hybrides compatibles avec les équations différentielles : sous quelles conditions une trajectoire d'un modèle hybride correspond à une trajectoire d'un système d'équations différentielles linéaires par morceaux. Le fait que les vitesses soient constantes dans un état qualitatif peut poser des problèmes de compatibilité entre les deux formalismes. Par exemple, dans le cas d'une spirale convergente en équations différentielles, plus la trajectoire se rapproche du centre de la spirale, plus la vitesse des variables ralentit. Que ce soit loin ou proche du centre, le système met le même temps pour faire un tour. Dans notre formalisme hybride, les vitesses constantes font que, plus la trajectoire se rapproche du centre, plus sa vitesse angulaire est grande.

Pour finir, le cadre de modélisation hybride est tout à fait adapté au domaine de la chronothérapie. Dans le cas de thérapies du cancer, le moment d'administration d'un médicament peut jouer un rôle très important dans son efficacité. En effet, contrairement à la majorité des cellules saines, les cellules cancéreuses se divisent en continu, quelle que soit l'heure de la journée. Avec sa simplicité et sa puissance, le formalisme hybride devrait permettre de comprendre le mode d'action du médicament et prédire son comportement avec un minimum de connaissances biologiques. La construction d'un modèle du fonctionnement de ce médicament facilitera alors le choix du meilleur moment de la journée pour l'administrer, celui où les effets sont les plus faibles sur les cellules saines synchronisées.

Le formalisme hybride présenté dans cette thèse a été la source de travaux sur l'identification des célérités (présenté brièvement dans le chapitre 8). Malgré la simplicité de ce formalisme, le nombre de célérités reste important et une méthode d'identification formelle des paramètres est indispensable.

9.2.3 Réduction de modèles hybrides

Afin de construire un lien entre les deux parties de cette thèse (réduction de réseau et modélisation hybride), nous pouvons imaginer une extension de la méthode de réduction adaptée au formalisme hybride. L'idée serait de supprimer un seuil d'une variable et adapter les célérités restantes afin de conserver les comportements asymptotiques du modèle hybride. Cependant, le formalisme hybride prenant en compte le temps, la notion de rapidité de la variable réduite présentée dans le chapitre 2 n'a plus beaucoup de sens. Il faudra considérer que les transitions de la variable autour du seuil sont prioritaires vis-à-vis des autres transitions, qui seront alors figées pendant ce temps.

Comme les signes des célérités d'un modèle sont issus des paramètres de Thomas, une partie de la réduction reposera sur le même principe que celle présentée dans cette thèse. Pour les valeurs de célérités, c'est plus difficile. En effet, le repliement des célérités peut poser des problèmes dans la conservation des comportements asymptotiques. De manière générale, l'aspect continu du formalisme rendra la construction des définitions à minima

très lourdes : pour conserver les comportements asymptotiques, il serait nécessaire de conserver également les délais du système et les positions de la trajectoire dans chacun des états discrets (parties fractionnaires).

De plus, ces comportements asymptotiques ne sont pas définis aussi simplement que dans le formalisme de Thomas. Dans un état stable, les célérités de toutes les variables ne sont pas forcément égales à 0 : par exemple, lorsqu'une trajectoire est « coincée » dans un coin de l'espace des états hybrides et qu'elle ne peut plus évoluer, elle atteint un état stable. Dans le cas d'un cycle, de la même façon que le modèle de l'horloge circadienne présenté dans le chapitre 7, certaines célérités pointent vers l'extérieur de l'espace des états, mais suivant la définition 19, les trajectoires ne sortent ni de l'espace des états ni du cycle (grâce aux glissements).

Le cas des hypercoins brièvement évoqué dans le chapitre 5 pourrait devenir une difficulté plus importante. En effet, la réduction de variables dans un réseau peut mener à l'apparition d'hypercoins augmentant la fréquence d'apparition de ce comportement.

Comme nous l'avons vu dans cette thèse, les méthodes formelles aident à la conception de modèles de systèmes biologiques, que ce soit des modèles discrets ou hybrides. Nous pensons que la généralisation des méthodes formelles ouvre la voie au développement de méthodes assistées par ordinateur de construction de modèles dans lesquelles les connaissances biologiques alimentent le modèle et le modèle suggère des expériences biologiques pertinentes. Dans cette optique, que ce soit la réduction de modèles ou l'extension hybride proposées dans cette thèse, ces deux sujets facilitent la conception de modèles de petite taille, et par la même, l'application systématique de méthodes formelles.

Bibliographie

- [1] G. K. Ackers, A. D. Johnson, and M. A. Shea. Quantitative model for gene regulation by lambda phage repressor. *Proceedings of the National Academy of Sciences*, 79(4) :1129–1133, 1982.
- [2] J. Ahmad, G. Bernot, J.-P. Comet, D. Lime, and O. Roux. Hybrid modelling and dynamical analysis of gene regulatory networks with delays. *ComPlexUs*, 3(4) :231–251, 2006 (Cover Date : November 2007).
- [3] J. Ahmad, O. Roux, G. Bernot, J.-P. Comet, and A. Richard. Analysing formal models of genetic regulatory networks with delays : Applications to lambda phage and T-cell activation systems. *International Journal of Bioinformatics Research and Applications*, 4(3) :240–262, 2008.
- [4] R. Alur, C. Belta, F. Ivančić, V. Kumar, M. Mintz, G. Pappas, H. Rubin, and J. Schug. Hybrid modeling and simulation of biomolecular networks. In M. Di Benedetto and A. Sangiovanni-Vincentelli, editors, *Hybrid Systems : Computation and Control*, volume 2034 of *Lecture Notes in Computer Science*, pages 19–32. Springer Berlin Heidelberg, 2001.
- [5] R. Alur, C. Courcoubetis, N. Halbwachs, T. A. Henzinger, P.-H. Ho, X. Nicollin, A. Olivero, J. Sifakis, and S. Yovine. The algorithmic analysis of hybrid systems. *Theoretical computer science*, 138(1) :3–34, 1995.
- [6] A. Arkin, J. Ross, and H. H. McAdams. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ -infected escherichia coli cells. *Genetics*, 149(4) :1633–1648, 1998.
- [7] S. L. Bailey and M. M. Heitkemper. Circadian rhythmicity of cortisol and body temperature : morningness-eveningness effects. *Chronobiology international*, 18(2) :249–261, 2001.
- [8] S. Becker-Weimann, J. Wolf, H. Herzl, and A. Kramer. Modeling feedback loops of the mammalian circadian oscillator. *Biophysical Journal*, 87(5) :3023 – 3034, 2004.
- [9] J. Behaegel, J. Comet, and M. Folschette. A hybrid hoare logic for gene network models. *arXiv : 1610.06715*, 2016.
- [10] J. Behaegel, J.-P. Comet, G. Bernot, E. Cornillon, and F. Delaunay. A hybrid model of cell cycle in mammals. In *6th International Conference on Computational*

- Systems-Biology and Bioinformatics (CSBio'2015)*, Bangkok (Thailand), November 22-25 2015.
- [11] J. Behaegel, J.-P. Comet, G. Bernot, E. Cornillon, and F. Delaunay. A hybrid model of cell cycle in mammals. *Journal of Bioinformatics and Computational Biology*, 14(1) :1640001 [17 pp.], 2016.
- [12] W. J. Belden, J. J. Loros, and J. C. Dunlap. Clock leaves its mark on histones. *Trends in Biochemical Sciences*, 31(11) :610–613, 2006.
- [13] G. Bernot and J.-P. Comet. On the use of temporal formal logic to model gene regulatory networks. In *International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics*, pages 112–138. Springer, 2009.
- [14] G. Bernot, J.-P. Comet, Z. Khalis, A. Richard, and O. Roux. A genetically modified Hoare logic. *arXiv : 1506.05887*, June 2015.
- [15] G. Bernot, J.-P. Comet, A. Richard, and J. Guespin. Application of formal methods to biological regulatory networks : Extending Thomas' asynchronous logical approach with temporal logic. *Journal of Theoretical Biology*, 229(3) :339–347, 2004.
- [16] G. Bernot, J.-P. Comet, and O. Roux. A genetically modified Hoare logic that identifies the parameters of a gene networks. In O. Roux and J. Bourdon, editors, *Proceedings of the 13th International Conference on Computational Methods in Systems Biology (CMSB)*, volume 9308 of *LNBI*, pages 8–12, Nantes, France, Sept. 16–18 2015.
- [17] G. Bernot, J.-P. Comet, and O. Roux. A genetically modified Hoare logic that identifies the parameters of a gene networks. In *Proceedings of the 13th Intl. Conf. on Computational Methods in Systems Biology (CMSB)*, volume 9308 of *LNBI*, pages 8–12, Nantes, France, Sept. 16–18 2015.
- [18] K. Bozek, A. Relógio, S. M. Kielbasa, M. Heine, C. Dame, A. Kramer, and H. Herzl. Regulation of clock-controlled genes in mammals. *PLoS One*, 4(3) :e4882, 2009.
- [19] T. Chen, H. L. He, G. M. Church, et al. Modeling gene expression with differential equations. In *Pacific symposium on biocomputing*, volume 4, page 4, 1999.
- [20] A. Cimatti, E. Clarke, E. Giunchiglia, F. Giunchiglia, M. Pistore, M. Roveri, R. Sebastiani, and A. Tacchella. Nusmv 2 : An opensource tool for symbolic model checking. In *International Conference on Computer Aided Verification*, pages 359–364. Springer, 2002.
- [21] J.-P. Comet, G. Bernot, A. Das, F. Diener, C. Massot, and A. Cessieux. Simplified models for the mammalian circadian clock. *Procedia Computer Science*, 11(0) :127 – 138, 2012. Proceedings of the 3rd International Conference on Computational Systems-Biology and Bioinformatics.
- [22] J.-P. Comet, J. Fromentin, G. Bernot, and O. Roux. A formal model for gene regulatory networks with time delays. In *1st International Conference on Computational Systems-Biology and Bioinformatics*, pages 1–13. Springer, 2010.

- [23] E. Cornillon, J.-P. Comet, G. Bernot, and G. Enée. *Proc. of the Thematic Research School on Advances in Systems and Synthetic Biology, Modelling Complex Biological Systems in the Context of Genomics*, chapter Hybrid Gene Networks : a new Framework and a Software Environment, pages 57–84. EDP Sciences, ISBN : 978-2-7598-1971-3, 2016.
- [24] M. I. Davidich and S. Bornholdt. Boolean network model predicts cell cycle sequence of fission yeast. *PloS one*, 3(2) :e1672, 2008.
- [25] H. De Jong, J.-L. Gouzé, C. Hernandez, M. Page, T. Sari, and J. Geiselman. Qualitative simulation of genetic regulatory networks using piecewise-linear models. *Bulletin of mathematical biology*, 66(2) :301–340, 2004.
- [26] J. Demongeot, A. Elena, and S. Sené. Robustness in regulatory networks : a multidisciplinary approach. *Acta Biotheoretica*, 56(1-2) :27–49, 2008.
- [27] E. W. Dijkstra. Guarded commands, nondeterminacy and formal derivation of programs. *Communications of the ACM*, 18(8) :453–457, 1975.
- [28] H. Echols. Bacteriophage λ development : temporal switches and the choice of lysis or lysogeny. *Trends in Genetics*, 2(0) :26 – 30, 1986.
- [29] H. Eisen, P. Brachet, L. P. d. Silva, and F. Jacob. Regulation of repressor expression in λ . *Proceedings of the National Academy of Sciences*, 66(3) :855–862, 1970.
- [30] H. El Snoussi and R. Thomas. Logical identification of all steady states : the concept of feedback loop characteristic states. *Bulletin of Mathematical Biology*, 55(5) :973–991, 1993.
- [31] E. Fanchon and A. Richard. Reduction of boolean networks. *Notes personnelles*, 2012.
- [32] A. Fauré, A. Naldi, C. Chaouiya, and D. Thieffry. Dynamical analysis of a generic boolean model for the control of the mammalian cell cycle. *Bioinformatics*, 22(14) :e124–e131, 2006.
- [33] D. B. Forger and C. S. Peskin. A detailed predictive model of the mammalian circadian clock. *PNAS*, 100(25) :14806–14811, 2003.
- [34] C. Gérard and A. Goldbeter. Temporal self-organization of the cyclin/cdk network driving the mammalian cell cycle. *PNAS*, 106(51) :21643–21648, 2009.
- [35] R. Ghosh and C. Tomlin. Symbolic reachable set computation of piecewise affine hybrid automata and its application to biological modelling : Delta-notch protein signalling. *Systems Biology*, 1(1) :170–183, 2004.
- [36] R. Ghosh and C. J. Tomlin. Lateral inhibition through delta-notch signaling : A piecewise affine hybrid model. In *International Workshop on Hybrid Systems : Computation and Control*, pages 232–246. Springer, 2001.
- [37] A. Goldbeter. Computational approaches to cellular rhythms. *Nature*, 420(6912) :238–245, 2002.

- [38] F. Guillaumond, H. Dardente, V. Giguère, and N. Cermakian. Differential control of small circadian transcription by rev-erb and ror nuclear receptors. *Journal of biological rhythms*, 20(5) :391–403, 2005.
- [39] K. R. Heidtke and S. Schulze-Kremer. Design and implementation of a qualitative simulation model of lambda phage infection. *Bioinformatics*, 14(1) :81–91, 1998.
- [40] C. A. R. Hoare. An axiomatic basis for computer programming. *Communications of the ACM*, 12(10) :576–580, 1969.
- [41] F. Jacob and J. Monod. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of molecular biology*, 3(3) :318–356, 1961.
- [42] R. F. Johnson, R. Y. Moore, and L. P. Morin. Loss of entrainment and anatomical plasticity after lesions of the hamster retinohypothalamic tract. *Brain research*, 460(2) :297–313, 1988.
- [43] M. Kaufman, J. Urbain, and R. Thomas. Towards a logical analysis of the immune response. *Journal of theoretical biology*, 114(4) :527–561, 1985.
- [44] Z. Khalis, J.-P. Comet, A. Richard, and G. Bernot. The SMBioNet method for discovering models of gene regulatory networks. *Genes, Genomes and Genomics*, 3(1) :15–22, 2009.
- [45] R. Khoodeeram, G. Bernot, and J.-Y. Trosset. An ockham razor model of energy metabolism. *Proc. of the Thematic Research School on Advances in Systems and Synthetic Biology*, P. Amar, F. Képès and V. Norris Eds, EDP Science pub., p.81-101, ISBN : 978-2-7598-2116-7, 2017.
- [46] B. Kuipers. Qualitative reasoning : Modeling and simulation with incomplete knowledge. *Automatica*, 25(4) :571 – 585, 1989.
- [47] P. Lavie. Sleep-wake as a biological rhythm. *Annual review of psychology*, 52(1) :277–303, 2001.
- [48] J.-C. Leloup and A. Goldbeter. Toward a detailed computational model for the mammalian circadian clock. *PNAS*, 100(12) :7051–7056, 2003.
- [49] J.-C. Leloup and A. Goldbeter. Modelling the dual role of per phosphorylation and its effect on the period and phase of the mammalian circadian clock. *Systems Biology, IET*, 5(1) :44–49, 2011.
- [50] P. Lincoln and A. Tiwari. Symbolic systems biology : Hybrid modeling and analysis of biological networks. In *International Workshop on Hybrid Systems : Computation and Control*, pages 660–672. Springer, 2004.
- [51] L. Mendoza, D. Thieffry, and E. R. Alvarez-Buylla. Genetic control of flower morphogenesis in *Arabidopsis thaliana* : a logical analysis. *Bioinformatics*, 15(7) :593–606, 1999.
- [52] D. Mestivier, P.-Y. Boëlle, K. Pakdaman, A. Richard, J.-P. Comet, G. Hutzler, C. Kuttler, A. Iartseva, and F. Képès. Modeling of λ phage genetic switch. Tutorial

- pour l'école thématique "Modelling Complex Biological Systems in the Context of Genomics", 2005.
- [53] A. Naldi, E. Remy, D. Thieffry, and C. Chaouiya. A reduction of logical regulatory graphs preserving essential dynamical properties. *Computational Methods in Systems Biology*, 2009.
- [54] A. Naldi, E. Remy, D. Thieffry, and C. Chaouiya. Dynamically consistent reduction of logical regulatory graphs. *Theoretical Computer Science*, 412(21) :2207 – 2218, 2011. Selected Papers from the 7th International Conference on Computational Methods in Systems Biology 7th International Conference on Computational Methods in Systems Biology.
- [55] B. Novak and J. J. Tyson. A model for restriction point control of the mammalian cell cycle. *Journal of theoretical biology*, 230(4) :563–579, 2004.
- [56] M. Ptashne. *A genetic switch : Gene control and phage lambda*. Palo Alto, CA (US) ; Blackwell Scientific Publications, Jan 1986.
- [57] M. Ptashne. A genetic switch, phage lambda and higher organisms. *Cell and Blackwell Scientific, Cambridge, MA*, 1992.
- [58] M. Ptashne and N. Hopkins. The operators controlled by the λ phage repressor. *Proceedings of the National Academy of Sciences*, 60(4) :1282–1287, 1968.
- [59] A. Rubinstein, V. Gurevich, Z. Kasulin-Boneh, L. Pnueli, Y. Kassir, and R. Y. Pinter. Faithful modeling of transient expression and its application to elucidating negative feedback regulation. *Proceedings of the National Academy of Sciences*, 104(15) :6241–6246, 2007.
- [60] A. Rubinstein, O. Hazan, B. Chor, R. Y. Pinter, and Y. Kassir. The effective application of a discrete transition model to explore cell-cycle regulation in yeast. *BMC research notes*, 6(1) :311, 2013.
- [61] H. Siebert and A. Bockmayr. Context sensitivity in logical modeling with time delays. In *International Conference on Computational Methods in Systems Biology*, pages 64–79. Springer, 2007.
- [62] H. Siebert and A. Bockmayr. Temporal constraints in the logical analysis of regulatory networks. *Theoretical Computer Science*, 391(3) :258 – 275, 2008.
- [63] R. Singhanian, R. M. Sramkoski, J. W. Jacobberger, and J. J. Tyson. A hybrid model of mammalian cell cycle regulation. *PLoS Comput Biol*, 7(2) :e1001077, 2011.
- [64] E. H. Snoussi. Qualitative dynamics of piecewise-linear differential equations : a discrete mapping approach. *Dynamics and stability of Systems*, 4(3-4) :565–583, 1989.
- [65] M. Sudo, K. Sasahara, T. Moriya, M. Akiyama, T. Hamada, and S. Shibata. Constant light housing attenuates circadian rhythms of mper2 mrna {AND} mper2 protein expression in the suprachiasmatic nucleus of mice. *Neuroscience*, 121(2) :493 – 499, 2003.

- [66] J. S. Takahashi, H.-K. Hong, C. H. Ko, and E. L. McDearmon. The genetics of mammalian circadian order and disorder : implications for physiology and disease. *Nature Reviews Genetics*, 9(10) :764–775, 2008.
- [67] D. Thieffry and R. Thomas. Dynamical behaviour of biological regulatory networks—II. immunity control in bacteriophage lambda. *Bulletin of Mathematical Biology*, 57(2) :277–297, 1995.
- [68] R. Thomas. Regulation of gene expression in bacteriophage lambda. In *Current Topics in Microbiology and Immunology/Ergebnisse der Mikrobiologie und Immunitätsforschung*, pages 13–42. Springer, 1971.
- [69] R. Thomas. Boolean formalization of genetic control circuits. *Journal of Theoretical Biology*, 42(3) :563 – 585, 1973.
- [70] R. Thomas. Logical analysis of systems comprising feedback loops. *Journal of Theoretical Biology*, 73(4) :631–656, 1978.
- [71] R. Thomas. Regulatory networks seen as asynchronous automata : A logical description. *Journal of Theoretical Biology*, 153(1) :1 – 23, 1991.
- [72] T. Tian, K. Burrage, P. M. Burrage, and M. Carletti. Stochastic delay differential equations for genetic regulatory networks. *Journal of Computational and Applied Mathematics*, 205(2) :696–707, 2007.
- [73] F. W. Turek, C. Joshu, A. Kohsaka, E. Lin, G. Ivanova, E. McDearmon, A. Laposky, S. Losee-Olson, A. Easton, D. R. Jensen, et al. Obesity and metabolic syndrome in circadian clock mutant mice. *Science*, 308(5724) :1043–1045, 2005.
- [74] J. J. Tyson and B. Novak. Regulation of the eukaryotic cell cycle : molecular antagonism, hysteresis, and irreversible transitions. *Journal of Theoretical Biology*, 210(2) :249–263, 2001.
- [75] J. J. Tyson and B. Novak. Temporal organization of the cell cycle. *Current Biology*, 18(17) :R759 – R768, 2008.
- [76] K. Vanselow and A. Kramer. Role of phosphorylation in the mammalian circadian clock. In *Cold Spring Harbor symposia on quantitative biology*, volume 72, pages 167–176. Cold Spring Harbor Laboratory Press, 2007.
- [77] U. Wilensky. Netlogo. <http://ccl.northwestern.edu/netlogo/>, Center for Connected Learning and Computer-Based Modeling, Northwestern University. Evanston, IL, 1999.
- [78] S. Yamaguchi, H. Isejima, T. Matsuo, R. Okura, K. Yagita, M. Kobayashi, and H. Okamura. Synchronization of cellular clocks in the suprachiasmatic nucleus. *Science*, 302(5649) :1408–1412, 2003.
- [79] R. Zhang, N. F. Lahens, H. I. Ballance, M. E. Hughes, and J. B. Hogenesch. A circadian gene expression atlas in mammals : implications for biology and medicine. *Proceedings of the National Academy of Sciences*, 111(45) :16219–16224, 2014.