

THESE



présentée à

L'UNIVERSITE DES SCIENCES ET TECHNOLOGIES DE LILLE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE

en

**PRODUCTIQUE,
AUTOMATIQUE et INFORMATIQUE INDUSTRIELLE**

par

Harald OHL

Ingénieur IDN/Ecole Centrale de Lille

**FONCTIONNEMENTS REPETITIFS DE SYSTEMES
FLEXIBLES DE PRODUCTION MANUFACTURIERE:
ANALYSE ET OPTIMISATION DES PERFORMANCES
A L'AIDE DE RESEAUX DE PETRI**

soutenue le 21 septembre 1995 devant la commission d'examen :

M. G. SCHMIDT	Président du jury	Professeur à l'Université de Munich
M. M. SILVA	Rapporteur	Professeur à l'Université de Saragosse
M. Y. FREIN	Rapporteur	Professeur à INP de Grenoble
M. Y. DALLERY	Rapporteur	Chargé de Recherche au CNRS - HDR
M. J.C. GENTINA	Directeur de thèse	Professeur à l'Ecole Centrale de Lille
M. J.P. BOUREY	Examinateur	Professeur à l'Ecole Centrale de Lille
M. J.P. CASSAR	Examinateur	MdC à l'Université de Lille I - HDR
M. E. CASTELAIN	Examinateur	MdC à l'Ecole Centrale de Lille

Thèse préparée au Laboratoire d'Automatique et Informatique industrielle de Lille
LAIL URA CNRS D1440 - Ecole Centrale de Lille (anciennement IDN).

Remerciements

Le travail présenté dans ce mémoire a été effectué au sein du Laboratoire d'Automatique et Informatique industrielle de Lille, LAIL URA CNRS 1440, à l'Ecole Centrale de Lille. Je tiens à remercier vivement Mme Geneviève Dauphin, Professeur à l'Ecole Centrale de Lille et Directeur du LAIL, de m'avoir accueilli au laboratoire. Le travail a été encadré par M. Jean-Claude Gentina, Professeur et Directeur de l'Ecole Centrale de Lille. Il m'a accordé un soutien efficace et chaleureux tout au long du travail. Je voudrais exprimer ici ma profonde gratitude.

Je voudrais particulièrement remercier

M. Manuel Silva, Professeur à l'Universidad de Zaragoza,
Departamento de Ingenieria Electrica e Informatica.

M. Yannick Frein, Professeur à l'Institut National Polytechnique de Grenoble,
Ecole de Génie Industriel et Laboratoire d'Automatique de Grenoble.

M. Yves Dallery, Chargé de Recherche au CNRS, Université de Paris VI,
Habilité à Diriger des Recherches, Laboratoire de Méthodologie et
Architecture des Systèmes Informatiques.

d'avoir accepté d'être rapporteurs de ma thèse ainsi que

M. Günther Schmidt, Professeur à la Technische Universität München,
Lehrstuhl für Steuerungs- und Regelungstechnik.

M. Jean-Pierre Bourey, Professeur à l'Ecole Centrale de Lille,
Laboratoire d'Automatique et Informatique industrielle de Lille.

M. Jean-Philippe Cassar, Maître de conférences à l'Université de Lille, Habilité à Diriger des Re-
cherches, Laboratoire d'Automatique et Informatique industrielle de Lille.

M. Emmanuel Castelain, Maître de conférences à l'Ecole Centrale de Lille,
Laboratoire d'Automatique et Informatique industrielle de Lille.

d'avoir accepté d'être examinateurs pour le jury. Je voudrais également remercier tous les membres du LAIL que je n'ai pas mentionnés auparavant et en particulier M. Hervé Camus pour le soutien qu'il m'a apporté et pour l'ambiance chaleureuse et conviviale qui régnait dans notre bureau. Enfin, je remercie M. Vangreveninge qui a assuré la reprographie de ce document.

Sommaire

Chapitre I: Introduction générale.

I.1 Les systèmes de production flexibles.....	5
I.2 Le projet CASPAIM.....	7
I.3 Objectifs et principales hypothèses de travail.....	9
I.4 Plan de la thèse.....	11

Chapitre II: Analyse des performances et conduite d'atelier.

II.1 Introduction.....	13
II.2 La gestion d'atelier par ratios de production.....	15
II.2.1 L'approche hiérarchisée de la gestion de production.....	15
II.2.2 Planification fine pour une production par ratio.....	17
II.2.3 Conduite optimale de flux en présence de pannes.....	21
II.2.4 Les approches de regroupement.....	27
II.3 Modèles et techniques pour l'évaluation des performances des systèmes à événements discrets.....	34
II.3.1 Prise en compte des événements discrets.....	34
II.3.2 La simulation.....	35
II.3.3 Les approches analytiques.....	37
II.4 Les Réseaux de Petri.....	43
II.4.1 Définitions et notations.....	43
II.4.2 Propriétés des RdP et caractérisation structurelle.....	45
II.4.3 Temporalisation des RdP.....	47
II.4.4 Les graphes d'événements.....	49
II.4.5 Modélisation de systèmes flexibles avec des graphes d'événements.....	50
II.5 Conclusion.....	53

Chapitre III: Fonctionnement répétitif d'un atelier flexible: Modélisation par Réseaux de Petri et optimisation des flux de production

III.1 Introduction.....	55
III.2 Fonctionnements répétitifs.....	57
III.2.1 Indices de performances.....	57
III.2.2 Classes de décisions.....	61

III.2.3 Formulation du problème	64
III.3 Construction du modèle Réseaux de Petri.....	66
III.3.1 Gammes de fabrication et ressources de transformation	66
III.3.2 En-cours et système de transport.....	70
III.3.3 Bouclage du modèle	76
III.3.4 Caractérisation structurelle.....	80
III.4 Evaluation préliminaire	82
III.4.1 Mise en équation	82
III.4.2 Détermination du routage	91
III.4.3 Exemple d'application.....	97
III.5 Conclusion.....	104

Chapitre IV: Modélisation et élaboration d'un ordonnancement pour un fonctionnement répétitif

IV.1 Introduction.....	105
IV.2 Construction d'un modèle ordonnançable.....	106
IV.2.1 Extraction des chemins de gamme et partition des charges	106
IV.2.2 Parallélisation et séquentialisation des gammes	111
IV.2.3 Contraintes pour un fonctionnement cyclique	122
IV.3 Ordonnancement et en-cours.....	125
IV.3.1 Position du problème.....	125
IV.3.2 Algorithme d'ordonnancement	126
IV.3.3 Description de l'algorithme.....	128
IV.3.4 Mise en œuvre.....	134
IV.4 Simplification du graphe d'événements	136
IV.5 Conclusion.....	138

V. Conclusion générale

V.1 Analyse des résultats.....	139
V.2 Perspectives de recherche	141

VI. Références bibliographiques

I. Introduction Générale

I.1 Les systèmes de production flexibles

Avec le développement de moyens de calcul de plus en plus puissants, la rationalisation des systèmes de production s'est orientée d'une mécanisation, généralement rigide, vers une automatisation plus souple, grâce notamment à l'introduction du concept de flexibilité. Ce concept dépasse largement le niveau de l'atelier de fabrication et peut, au plus haut niveau décisionnel d'une entreprise, avoir une interprétation purement économique, liée au contexte concurrentiel dans lequel se trouve l'entreprise. Le pouvoir d'adaptation en termes statiques, tels que volume de production ou nature des produits, ainsi qu'en termes dynamiques, tels que temps de réponse à une demande fluctuante ou temps d'industrialisation d'un nouveau produit est d'une importance cruciale pour la compétitivité d'une entreprise. La réalisation effective de ces objectifs de nature économique repose néanmoins sur les niveaux inférieurs et se traduit au niveau de l'atelier de fabrication par des besoins de flexibilité à caractère technique.

La notion de flexibilité est liée à celle d'incertitude car le maintien d'une flexibilité ne se justifie pas dans un environnement entièrement déterministe. En l'absence d'incertitudes il est préférable de déterminer, une fois pour toutes, le meilleur choix, et de l'implanter définitivement dans le système étudié. Cette considération est valable aussi bien au niveau de la conception d'un système de production qu'au niveau de son exploitation. Si une demande est parfaitement connue et stable sur un très long horizon, rien n'est en général plus rentable que de concevoir des lignes de production dédiées avec des ressources dédiées. En revanche, pour une fabrication de type atelier où des produits sont fabriqués à l'unité, des machines universelles qui permettent de traiter un grand nombre de produits différents sont plus adaptées.

Les ateliers flexibles se situent entre ces deux situations extrêmes et constituent un compromis entre la productivité et la flexibilité, deux critères habituellement antinomiques. Un atelier flexible permet la production économique de petites ou moyennes séries pour suivre une demande fluctuante, sans pour autant devoir gérer d'importants stocks de produits finis. Les ressources qui constituent un atelier flexible doivent donc être capables d'effectuer des opérations différentes sur des produits différents. Si les

temps de changement d'outil entre des opérations sur des produits différents sont petits, un atelier flexible permet alors la fabrication simultanée de plusieurs types de produits.

Un autre type de flexibilité concerne le niveau opérationnel et l'exploitation d'un atelier. Avec des machines fiables et en l'absence d'autres sources de perturbations, l'élaboration d'un ordonnancement prévisionnel, c'est-à-dire calculé hors ligne et figé sur un horizon déterminé, permet de prendre en compte un critère d'optimisation global. En revanche, si le système est soumis à des perturbations, la possibilité de résoudre des conflits le plus tard possible peut permettre d'atténuer les conséquences d'aléas, comme des pannes de machines par exemple.

Partant du principe que la commande d'un système de production doit être déterministe, toute flexibilité doit être tôt ou tard résolue. L'élaboration de la commande d'un système de production se complique alors très vite avec des degrés de flexibilité croissants. Plusieurs types de problèmes se posent, principalement liés à la nature discrète des systèmes considérés et à l'explosion combinatoire du nombre d'états que peuvent prendre ces systèmes.

Le premier type de problèmes concerne la vérification de certaines propriétés qualitatives que la commande du système doit posséder. Les propriétés généralement requises sont l'absence de blocages, la possibilité de revenir à tout moment à un état défini ainsi que le respect des contraintes spécifiques au système physique considéré. Ce dernier point concerne par exemple la prise en compte des capacités des emplacements de stockage.

Un deuxième type de problèmes se pose lors de l'analyse des performances du système. La commande peut utiliser les degrés de liberté mis à sa disposition pour optimiser des critères de performance. Il faut alors déterminer ces critères ainsi que la commande qui les optimise. Avec le développement des systèmes flexibles, de nouvelles classes de problèmes d'ordonnancement sont apparues, généralement très difficiles à résoudre et souvent NP-complets [CAR 88][GOT 93].

Une troisième problématique relève des questions d'implantation physique de la commande. En particulier la commande de bas niveau est souvent répartie sur des architectures matérielles différentes et géographiquement distantes. Ceci pose divers problèmes de communication, de synchronisation et de mise à jour d'informations réparties.

De plus, il faut assurer que le code informatique effectivement implanté dans les moyens physiques corresponde bien à la commande élaborée.

Les problèmes évoqués ont fait l'objet de recherches intensives, mais, bien que de nombreux résultats soient disponibles, peu de réalisations industrielles existent. Ceci est d'autant plus étonnant que les avantages des systèmes flexibles sont indiscutables, surtout dans un contexte de concurrence accrue. La tergiversation des industriels peut s'expliquer en partie par une certaine méfiance vis-à-vis d'une euphorie du 'tout-automatisé' et un retour vers une automatisation limitée et moins coûteuse qui répond mieux au besoin de rentabilité financière à court terme.

Un autre facteur critique est la nécessité de proposer des méthodes de conception adaptées à la complexité des systèmes flexibles. Afin de rendre les résultats de recherche - parfois très ponctuels - effectivement utilisables et accessibles, ils doivent être structurés et regroupés en méthodes. Ces méthodes doivent intégrer les différents aspects mentionnés auparavant et fournir une assistance aussi bien au niveau de la conception qu'au niveau de l'exploitation d'un système flexible.

I.2 Le projet CASPAIM

La problématique décrite dans le paragraphe précédent constitue le thème central du projet CASPAIM, acronyme de *Conception Assistée de Systèmes de Production Automatisés dans l'Industrie Manufacturière*, développé au Laboratoire d'Automatique et d'Informatique industrielle à l'Ecole Centrale de Lille. Bien que le but de ce projet soit la mise au point d'un ensemble complet de méthodes de conception, son développement s'est limité dans un premier temps à l'aspect commande, laissant de côté les questions qui relèvent de la conception du système physique.

Dès le début du projet, les Réseaux de Petri ont constitué le principal outil de modélisation. Dans la première version, baptisée CASPAIM I, le point de départ de la construction du modèle était le *prégraphe*, modélisé par un Réseau de Petri coloré. Cette approche de modélisation a été développée dans [KAP 88] et s'inspirait de l'architecture physique pour décrire le cheminement possible des différents types de produits dans le système. Dans le modèle résultant, une place correspondait à un lieu physique, une transition à une opération de transformation ou transfert et une marque colorée à un pro-

duit d'un certain type. Pour raffiner la modélisation, des primitives de structuration ont été développées dans [BOU 88]. Après l'intégration de ces modules, explicitant les traitements et les transferts dans le prégraphe, un modèle complet, appelé graphe de commande, a été obtenu. La vérification des propriétés qualitatives et l'analyse quantitative du comportement s'effectuaient ensuite sur ce modèle par simulation [CAS 87]. D'autres travaux dans le cadre de CASPAIM I abordaient les questions d'implantation de la commande, notamment sur des architectures réparties [CRA 89].

Lors de l'application à des exemples réels, certains inconvénients de la méthode sont apparus. Le modèle ne possédait qu'un seul niveau d'abstraction, ce qui rendait difficile une représentation fine du comportement de ressources ayant une structure dynamique, comme des stockeurs rotatifs ou des convoyeurs par exemple. Ceci provenait du fait qu'il n'existait pas de modèle spécifique du procédé. En effet, les seules informations relatives aux ressources physiques étaient intégrées dans le graphe de commande et représentaient uniquement la capacité des zones de stockage ainsi qu'une temporisation des opérations pour une évaluation des performances par simulation.

Par ailleurs, l'interfaçage avec la partie décisionnelle, qui se trouvait dans un niveau hiérarchique à l'extérieur du modèle Réseaux de Petri, faisait appel à des arcs adaptatifs [GEN 88]. L'utilisation de ces arcs adaptatifs permet certes une modélisation fine et compacte, mais implique la perte de certaines propriétés formelles des Réseaux de Petri et interdit en particulier toute analyse structurelle basée sur les invariants. Ces inconvénients ont motivé le développement d'une nouvelle approche de modélisation dont les éléments de base ont été proposés dans [CRU 91].

Cette nouvelle approche s'appuie sur la dualité entre les ressources et les produits, en les mettant au même niveau dans la modélisation. Pour les produits, la notion de gamme de fabrication est introduite, modélisant explicitement les contraintes de précedence entre les opérations à effectuer sur les produits. Comme chaque type de produit possède sa propre gamme, l'utilisation de jetons colorés n'est plus nécessaire pour distinguer des produits de types différents. Cette approche, appelée CASPAIM II, a été développée selon plusieurs axes, intégrant notamment des aspects de surveillance [TOG 92] et [ELK 93] et une gestion des modes de marche [BOI 91]. Dans [AMA 94] l'approche produit, qui était initialement centrée sur la partie logique, a été complétée par une méthode de modélisation systématique du procédé, basée sur une représentation structuro-fonctionnelle et utilisant un formalisme orienté objet. La structuration de la

partie commande en fonction des relations d'accessibilité a été l'objet de [AUS 94] abordant aussi le problème de la vérification formelle d'absence de blocages du modèle Réseaux de Petri. Dans [HUV 94] une démarche d'implantation a été proposée s'appuyant sur les concepts de généricité et de modularité utilisés en phase de conception. Cette démarche d'implantation fait appel au langage ADA qui permet une réalisation informatique aisée de ces concepts et facilite une implantation répartie de la commande ainsi que la simulation de celle-ci pour une analyse préalable de son comportement.

En ce qui concerne l'analyse quantitative, deux aspects particuliers ont été étudiés dans le cadre du projet CASPAIM II. Dans [HAM 91] une heuristique d'ordonnancement pour des ateliers de type job-shop a été proposée. Elle s'appuie principalement sur une décomposition temporelle du problème et a pour objectif la minimisation des temps d'attente et de transit. Le pilotage d'atelier, en tant que mise en œuvre d'un ordonnancement prévisionnel, a été étudié dans [TAW 95]. Dans ces travaux les possibilités de modifications locales d'un ordonnancement prévisionnel sont examinées dans le but de lui conférer un aspect de réactivité limitée.

I.3 Objectifs et principales hypothèses de travail

L'objectif des travaux présentés dans ce mémoire est de développer la composante d'analyse quantitative du projet CASPAIM pour y englober une évaluation préliminaire des performances basée sur un calcul de flux. Ensuite, en précisant davantage les critères d'optimisation et en raffinant le modèle utilisé, l'objectif de l'approche est d'élaborer progressivement une commande effectivement applicable à l'atelier. La perspective d'une intégration dans le projet CASPAIM nous a amené à considérer les deux points suivants:

Premièrement, nous insistons sur l'aspect méthodique et la structuration en étapes distinctes pour tenir compte du grand nombre d'hypothèses qui conditionnent, parfois implicitement, une analyse de propriétés quantitatives. De plus, une structuration de l'approche crée une ossature qui facilite de futures extensions en proposant un positionnement clair par rapport à l'existant.

Deuxièmement, nous avons intégré certains des acquis du projet CASPAIM notamment en ce qui concerne le choix de l'outil de modélisation. La plupart des travaux effectués dans le cadre du projet CASPAIM utilisent une modélisation basée sur des

Réseaux de Petri. La puissance de cet outil et son adéquation pour la modélisation et l'évaluation des systèmes de production ont été démontrées dans de nombreux cas [VAL 86][SIL 89][DIC 93][PRO 95]. La disponibilité d'un grand nombre de résultats établis avec cet outil a aussi facilité ce choix.

A l'intérieur de ce cadre très général - développement d'une méthode s'appuyant sur des Réseaux de Petri - nous avons fait un certain nombre d'hypothèses concernant le processus de fabrication, le mode de gestion de l'atelier et le système de transport.

Dans tout le mémoire nous nous limitons à l'étude de systèmes de production discrets. Ceci implique d'un côté que les produits sont identifiables et localisables individuellement et de l'autre côté que le processus de fabrication peut être caractérisé par des changements d'états discrets. Cette caractérisation est évidemment liée au niveau d'abstraction choisi pour la modélisation. En effet, l'usinage d'une pièce mécanique peut être perçu comme un processus continu, en regardant de près l'avancement d'une broche de fraisage. Ce même processus peut être modélisé par des états discrets et des événements provoquant des changements d'état, comme par exemple *début d'opération / opération en cours / fin d'opération*. En ce qui concerne les produits, la modélisation de pièces solides par un flux continu peut être une approximation adaptée pour certains calculs d'optimisation.

Une partie de l'analyse de performance présentée dans ce mémoire est basée sur l'hypothèse de la conservation des produits. Nous ne considérons donc pas des systèmes d'assemblages faisant intervenir des composants qui sont modélisés explicitement. En revanche, des systèmes dans lesquels l'assemblage peut être assimilé à une opération de transformation, rentrent dans le cadre de ce mémoire.

Il est courant de caractériser les ateliers de fabrication en utilisant une classification en *flow-shop*, *job-shop* ou *open-shop*. Dans des *flow-shops* l'ordre de visite des machines est le même pour tous les produits. Des représentants typiques de cette classe d'ateliers sont des lignes de montage dans l'industrie automobile par exemple. Dans les *job-shops* l'ordre de visite peut être différent suivant le type de produit. Une flexibilité extrême est atteinte dans les *open-shops*, où chaque produit peut visiter les machines dans un ordre quelconque. Ce dernier type de fabrication se retrouve par exemple dans des ateliers d'assemblage électronique où la structure physique des éléments, une platine recevant des composants, n'impose pas de contraintes de précédence.

Les systèmes que nous considérons dans ce mémoire sont des ateliers de fabrication mécanique et nous supposons que des contraintes de précédence existent entre la plupart des opérations. Nous admettons néanmoins certains types de flexibilités que nous spécifions lors de la construction de notre modèle. Au regard de la classification précédemment citée, ces systèmes peuvent alors être qualifiés de *job-shops avec flexibilités de gammes limitées*.

Un élément clef dans un atelier de fabrication automatisé est le système de transport. Il relie les ressources de transformation et détermine ainsi les relations d'accessibilité. Dans ce qui suit, nous supposons toujours que le système de transport est fortement connexe. En pratique, la réalisation physique de tels systèmes de transport peut être effectuée avec des chariots guidés automatiquement ou avec une structure de convoyeur. La première solution est coûteuse et techniquement plus difficile à réaliser. L'autonomie des chariots et le fait qu'ils circulent sur des voies qui ne leur sont généralement pas exclusivement réservées, engendrent de nouveaux problèmes, en particulier la possibilité de collisions. Dans des grandes installations avec des distances de transport importantes des chariots guidés automatiquement peuvent cependant être adaptés.

I.4 Plan de la thèse

Après ce premier chapitre d'introduction, suivent trois autres chapitres principaux et une conclusion.

Le deuxième chapitre détaille le cadre à l'intérieur duquel se situent nos travaux. D'abord nous introduisons la problématique de la gestion de production dans le contexte d'une approche par ratio de production et présentons deux approches de conduite de flux. Nous dégageons les caractéristiques de ces approches et les problèmes rencontrés lors de leurs applications à la problématique traitée. Puis nous passons en revue différentes approches de modélisation et méthodes d'analyse de systèmes à événements discrets et étudions leurs hypothèses et leur adéquation pour l'étude des problèmes évoqués. Nous justifions le choix des Réseaux de Petri et présentons une approche de modélisation de systèmes de production flexibles par graphes d'événements temporisés.

Dans le troisième chapitre nous abordons le problème d'élaboration d'une commande pour un fonctionnement répétitif par une analyse de flux. D'abord nous identifions les indices de performances pour un fonctionnement répétitif et analysons leurs interactions et les types de décisions qui sont en rapport avec leur optimisation. Ensuite nous présentons une méthode pour développer le modèle Réseaux de Petri qui accompagnera par des transformations successives le processus d'élaboration de la commande. Au cours de cette étape de modélisation, nous précisons alors davantage la classe des systèmes qui rentrent dans le cadre de ce mémoire. La deuxième partie de ce chapitre est consacrée à la mise en équation du problème d'optimisation des indices de performances retenus. Pendant cette phase d'optimisation nous tenons compte des contraintes dues à la nature discrète des systèmes considérés et établissons des conditions nécessaires et suffisantes pour une réalisation effective.

Le quatrième chapitre concerne le passage du niveau flux au niveau discret. Dans un premier temps, nous identifions les classes de décisions qui restent à résoudre et mettons en évidence les facteurs qui contribuent à la combinatoire du problème étudié. En particulier, nous introduisons une nouvelle approche pour transformer le modèle issu de l'analyse de flux en un modèle totalement déterministe. Nous comparons cette nouvelle approche à une approche existante et établissons à cet effet des bornes pour le dernier indice de performance à optimiser. Nous montrons que ces bornes peuvent être atteintes et soulignons ainsi l'intérêt de l'approche proposée. Une méthode de parcours de l'espace de recherche en découle et nous permet de réduire la complexité du problème traité. A la fin de ce chapitre nous présentons une heuristique d'ordonnancement qui transforme le modèle en un graphe d'événements.

Nous concluons ce mémoire en rappelant les apports des travaux présentés et en indiquant les principales perspectives de recherche à développer.

II. Analyse des performances et conduite d'atelier

II.1 Introduction

Une des difficultés rencontrées dans l'analyse des systèmes de production réside dans le fait que leurs performances dépendent autant de leurs caractéristiques physiques que de la manière de les exploiter. Ce très fort couplage entre le système physique et sa commande est une caractéristique des systèmes à événements discrets qui complique fortement leur analyse. Dans les systèmes de production, la commande se traduit à la fois par des décisions d'affectation dans le domaine spatial et des décisions de séquençement ou positionnement dans le domaine temporel. Les conséquences de chacune de ces décisions peuvent aussi bien être faibles, et même ne pas avoir d'influence du tout, ou au contraire avoir des répercussions importantes sur le comportement du système.

En toute rigueur, il faudrait alors résoudre tous les indéterminismes qui existent dans un système pour analyser avec précision tous les aspects de son comportement. Une résolution simultanée de tous les indéterminismes est pratiquement impossible pour des systèmes de tailles importantes en raison de l'explosion combinatoire du nombre de solutions envisageables, mais de façon plus réaliste à cause des incertitudes qui peuvent exister sur les données du problème. Dans certains cas, pour une conception préliminaire par exemple, une analyse approximative peut être suffisante. Si, par contre, l'objectif est d'effectuer une analyse précise et d'élaborer une commande, le seul moyen d'appréhender la complexité consiste à utiliser une décomposition en niveaux hiérarchiques. A chacun de ces niveaux hiérarchiques correspond alors un modèle adapté et des hypothèses spécifiques associées. Pour les systèmes de production que nous considérons, ces hypothèses concernent notamment l'horizon temporel, la finesse de la modélisation et la manière de traiter les conflits.

Une hypothèse simplificatrice couramment utilisée pour effectuer une première analyse consiste à négliger le caractère discret de ces systèmes et à adopter une modélisation par flux continu. Justifier cette simplification conduit à faire un certain nombre d'hypothèses supplémentaires, concernant notamment les fréquences d'apparition des événements qui ont lieu dans le système. La modélisation par flux continu

suppose alors que les machines traitent tous les produits simultanément dans un mode dit de *processeur partagé*, similaire à celui développé pour des applications multitâches en informatique. Ceci implique en particulier que les phénomènes d'attente dus à des indisponibilités de ressources et propres aux systèmes discrets, n'existent plus. Dans cette approche de modélisation, des éventuels partages de flux de produits sont représentés par de simples fractions qui remplacent les décisions de routage.

Un autre moyen de contourner la résolution explicite des indéterminismes consiste à employer une approche stochastique. Dans ces approches les conflits sont résolus de manière probabiliste ce qui permet de maintenir une modélisation sous forme de système discret. Malheureusement, les outils basés sur cette approche probabiliste requièrent des hypothèses restrictives et difficilement justifiables si une résolution analytique est envisagée. Ainsi, les Réseaux de Petri stochastiques et la théorie des files d'attente, approches probabilistes les plus répandues, nécessitent des temporisations de type exponentiel pour une résolution analytique. Cette hypothèse permet néanmoins d'aborder une complexité d'analyse qui, dans le cas contraire, ne pourrait être maîtrisée pour l'étude des systèmes réels. En présence d'autre lois de distribution qui ne peuvent pas se ramener à la loi exponentielle, la simulation reste alors le seul moyen d'analyser des performances.

Lors d'une analyse préliminaire se posent plusieurs types de problèmes. Premièrement il s'agit d'évaluer la nature des approximations pour savoir si les résultats évalués constituent éventuellement des majorants ou minorants des indices de performances, ou si ces approximations entraînent des écarts dans les deux sens. Dans tous les cas une estimation de l'erreur peut s'avérer utile. Le deuxième type de problème concerne la cohérence des résultats d'un niveau avec les contraintes spécifiques du niveau hiérarchique inférieur, en particulier lors du passage d'un calcul de flux à un niveau qui tient compte du caractère discret. Cette problématique présente des aspects particuliers pour une production par ratio, surtout lorsqu'une conduite par ordonnancement répétitif est envisagée.

Nous abordons l'étude de cette problématique par la présentation de quelques approches de planification pour une production par ratio et évoquons les problèmes relatifs à l'interaction avec le niveau ordonnancement. Nous passons ensuite en revue différents modèles et méthodes d'évaluation de performances utilisables pour la problématique considérée et justifions le choix des Réseaux de Petri. Enfin nous introduisons

formellement les Réseaux de Petri et présentons une approche de modélisation des ateliers flexibles par des graphes d'événements.

II.2 La gestion d'atelier par ratios de production

II.2.1 L'approche hiérarchisée de la gestion de production

Afin d'une part de maîtriser la complexité des problèmes rencontrés et d'autre part de tenir compte des incertitudes, la gestion de production est généralement abordée suivant une approche scindée en différents niveaux hiérarchiques [HAR 92]. Dans une telle décomposition hiérarchique, les horizons temporels deviennent de plus en plus courts et les données de plus en plus précises en descendant les niveaux. Aussi, les données issues d'un niveau supérieur sont traitées comme des contraintes par le niveau inférieur. Il peut arriver que ces contraintes soient trop dures de sorte qu'un niveau ne puisse pas trouver de solution compatible. Dans ce cas il faut remonter dans la hiérarchie et relaxer les contraintes au niveau supérieur. Ceci conduit alors naturellement à des approches dans lesquelles la cohérence entre les niveaux consécutifs est assurée par des techniques itératives*.

Une méthode de gestion de production répandue dans le milieu industriel et qui utilise les concepts mentionnés ci-dessus est MRP II (*Material Requirements Planning*). Dans un environnement de ce type, le plus haut niveau est celui de la planification sur un horizon long, 12 mois par exemple. Ce niveau tient compte des commandes fermes, des prévisions de ventes ainsi que des stocks existants et souhaités. Le résultat de cette étape est un *Plan Directeur de Production* (PDP) qui est ensuite validé par rapport aux ressources critiques. Pour établir ce plan, les nomenclatures des produits finis sont éclatées et les charges induites sur les ressources critiques sont calculées sur une période d'un mois par exemple. A ce stade des ajustements à long terme peuvent être entrepris pour équilibrer des variations saisonnières.

Ces ajustements peuvent concerner la demande, la capacité de production ou les deux à la fois. La demande est influençable à travers des actions commerciales, le prix

* voir [DAU 92] par exemple pour une étude de cohérence entre les niveaux de planification et d'ordonnement pour une fabrication de type atelier.

des produits, ou la renégociation des dates de livraison. En revanche, l'ajustement de la capacité peut se faire par la sous-traitance, l'extension du système de production ou une meilleure utilisation des ressources disponibles en travaillant en 3x8 par exemple. Cette problématique consistant à ajuster la capacité et la demande nécessite évidemment la prise en compte des niveaux de stocks. La tenue d'un stock à plus long terme peut également être envisagée bien que cette option soit généralement peu prisée.

La programmation se situe à un niveau inférieur à la planification. Elle est effectuée sur un horizon plus court (deux mois par exemple) et avec une période plus courte (une semaine par exemple). De la même manière comme au niveau supérieur, les nomenclatures des produits finis sont éclatées mais à ce niveau les charges induites sont calculées pour toutes les ressources du système. Là aussi, des ajustements peuvent être nécessaires, car des surcharges ponctuelles peuvent subsister, même si la charge moyenne, étalée sur le mois, est admissible*.

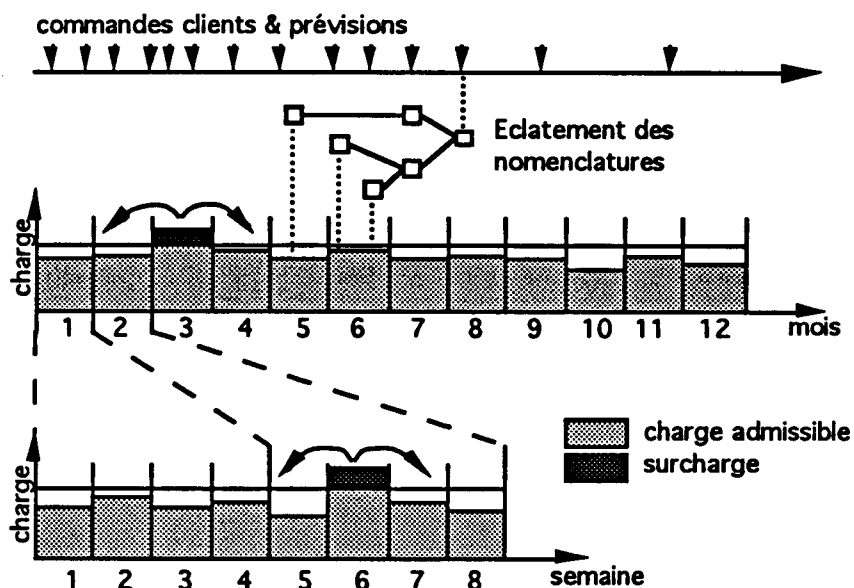


Fig. II-1 Planification à long terme et programmation

Le résultat de cette étape est un programme prévisionnel de production dans lequel les données sont figées sur un certain horizon. Ces données figées sont ensuite utilisées

* voir [MAZ 89] pour une formulation en problème de programmation avec des variables 0-1.

pour la génération des commandes d'approvisionnement et servent aussi de référence au niveau inférieur. Ce niveau inférieur peut être réalisé de plusieurs façons:

Dans la méthode MRP telle qu'elle est généralement implantée, s'enchaîne alors la détermination des *lots de fabrication* selon des règles de lotissement spécifiées. Ces règles peuvent tenir compte des contraintes technologiques qui influencent les tailles des lots, comme par exemple les temps de configuration des machines ou les tailles de conteneurs de transport. Des *ordres de fabrication* (OF) sont émis pour des lots de produits et ces OF sont ensuite ordonnancés. Etant donné que des systèmes gérés en MRP fonctionnent avec des stocks de produits finis, les dates dues des OF ne correspondent plus nécessairement aux dates dues des commandes des clients.

Une autre approche, plus adaptée aux cas des ateliers flexibles, consiste à employer une gestion par *ratios de production*. Une telle approche peut se justifier par le fait qu'un atelier flexible fortement automatisé est seulement rentable, si la demande pour chaque type de produit dépasse une certaine limite. En effet, les investissements nécessaires pour préparer un atelier flexible à la fabrication d'un nouveau type de produit peuvent être substantiels. Le coût de développement des supports de bridage et des préhenseurs pour les robots de manutention sont généralement élevés. Il faut donc être certain de produire des quantités suffisantes pour limiter le coût ramené à la pièce.

En revanche, une fois configuré, un système flexible permet de produire simultanément des types de produits différents et ceci avec des ratios de production ajustables. Cette possibilité constitue un degré de liberté important qui peut être utilisé pour l'optimisation d'un critère. Il se pose alors le problème de la détermination de ce critère et de la traduction des besoins de fabrication en ratios de production.

II.2.2 Planification fine pour une production par ratio

Nous supposons que les données issues du niveau programmation sont les quantités absolues (*besoins nets* en termes de MRP) des différents types de produits à fabriquer sur une période figée, par exemple d'une semaine. Au niveau de la planification fine une approche par flux est généralement utilisée afin d'éviter la complexité d'une approche discrète. Pour justifier cette simplification nous adoptons les hypothèses sur les

échelles temporelles des événements pris en compte, énoncées dans [KIM 83] et connues sous le nom d'*hypothèses de Gershwin*.

Ces hypothèses ne concernent pas seulement les échelles temporelles des événements, mais également d'autres aspects du processus comme par exemple les lois qui régissent les pannes des machines. Elles sont fréquemment retenues à cause de leur caractère réaliste, surtout dans le cas des ateliers flexibles avec des machines relativement fiables. Selon ces hypothèses, les fréquences d'occurrences des événements dans les quatre niveaux identifiés diffèrent d'au moins un ordre de grandeur. La notion de *période* caractérise ici un intervalle de temps moyen entre deux événements atomiques d'un niveau hiérarchique considéré:

- (1) La plus petite période est celle d'un changement (set-up) entre deux opérations pour lesquelles une machine a été préconfigurée.
- (2) La période suivante correspond à la durée d'une opération de transformation représentative.
- (3) La période suivante est celle entre deux pannes (MTTF, Mean Time To Failure) ou celle pour la réparation d'une machine (MTTR, Mean Time To Repair).
- (4) La plus grande période est celle pour laquelle les besoins de fabrication ont été spécifiés.

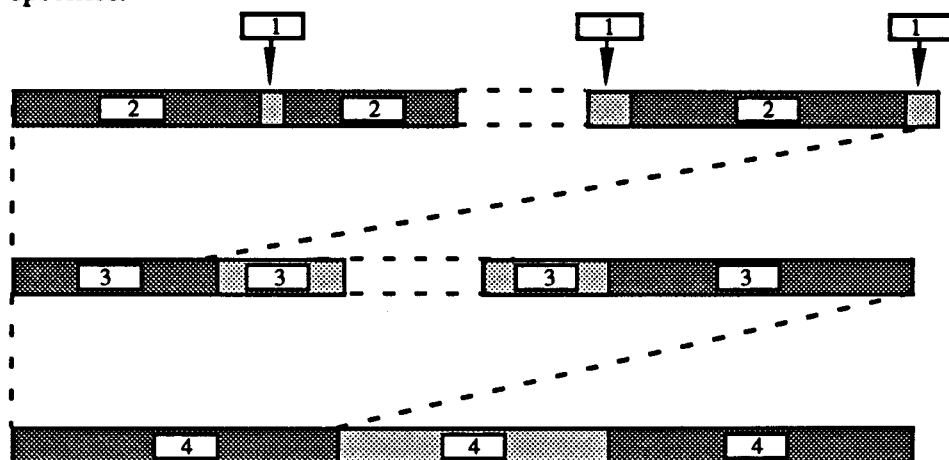


Fig. II-2 Echelles temporelles des différentes classes d'événements

Pour fixer les idées, des ordres de grandeur réalistes pour les quatre niveaux sont respectivement de quelques secondes, quelques minutes, quelques heures et quelques

jours ou une semaine. Les conséquences de ces hypothèses pour l'analyse sont importantes et vont bien au-delà d'une justification de l'approche par flux continu:

La différence de l'ordre de grandeur entre les périodes de (1) et (2) permet de négliger les temps de changement d'outil (set-up times) des machines pour une configuration donnée. En ce qui concerne l'ordre de grandeur des périodes de (2) et (3), l'hypothèse des pannes rares par rapport au nombre d'opérations permet de supposer l'existence d'un régime de production stable entre des événements qui caractérisent des débuts ou fins de pannes. Le fait de travailler sur une période de production avec des besoins stables (4), qui est grande devant l'ordre de grandeur des opérations (2), justifie une approche par flux continu et l'utilisation de ratios de production. Finalement, le grand nombre de pannes pendant la période de production (4) permet d'estimer la disponibilité moyenne des machines avec une certaine confiance.

Remarquons que les hypothèses de Gershwin n'évoquent pas les ordres de grandeur des temps de transport. Ceci peut s'expliquer par le fait que ces hypothèses ont été énoncées dans le cadre d'une approche de conduite de flux où les temps de transport sont généralement négligés. Les durées des temps de transport apparents* sont effectivement difficiles à estimer car elles ne dépendent pas seulement des caractéristiques physiques et de l'architecture d'un système, mais également des aspects liés à l'exploitation, comme par exemple le niveau de l'en-cours.

L'approche par flux continu implique de plus que les produits ne sont plus identifiables individuellement et par conséquent que la notion de date due n'existe plus à l'intérieur d'une période élémentaire. Nous supposons que la programmation a été effectuée d'une telle manière que la charge de travail programmée pour la période élémentaire considérée puisse effectivement être absorbée par l'atelier. Pour la mise en équation, nous utilisons les notations suivantes:

I	nombre de types de produits différents, $i = 1..I$
T	durée de la période élémentaire considérée
$D_i(T)$	demande en nombre de produits de type i , sur $[0..T]$
$D(T)$	demande totale, $D(T) = \sum D_i(T)$, sur $[0..T]$

* cette notion sera précisée dans le chapitre III.

\bar{d}_i	taux moyen de demande de produits de type i, $\bar{d}_i = D_i / T$
b_i	fraction de produits i de la demande totale, $b_i = D_i/D$
r_i	ratio de production pour les produits de type i, $\sum r_i = 1$
Φ	taux de production maximal en fonction des ratios r_i $\Phi = \Phi(r_1, r_2, \dots, r_I)$ mesuré en pièces par unité de temps
Φ_i	taux de production de produits de type i,

Une première manière d'aborder le problème consiste à ne pas considérer de pannes de machines et à prendre les ratios de production directement proportionnels aux besoins:

$$r_i = b_i \quad i = 1..I \quad (\text{II.1})$$

Les ratios r_i ainsi déterminés indiquent seulement la proportion relative de chaque type de produits mais ne spécifient pas les taux de production absolus. Pour fixer ces taux, deux possibilités simples sont concevables. La première consiste à étaler la production régulièrement sur la période T et nous obtenons des taux de production définis par

$$\Phi_i = \bar{d}_i = \frac{D_i(T)}{T} \quad i = 1..I \quad (\text{II.2})$$

La deuxième possibilité consiste à caler la production le plus possible à gauche, c'est-à-dire de produire à saturation pendant un temps T' . Comme l'atelier est par hypothèse capable d'absorber la charge de travail pendant la période T , nous avons $T' \leq T$. Dans ce cas les Φ_i sont définis par

$$\Phi_i = r_i \Phi(r_1, r_2, \dots, r_I) \quad i = 1..I \quad (\text{II.3})$$

et la durée T' par

$$T' = \frac{D(T)}{\Phi(r_1, r_2, \dots, r_I)} \quad (\text{II.4})$$

La quantité Φ qui intervient dans le calcul des Φ_i et de T' est a priori inconnue et ne peut pas être déterminée avec précision sans expliciter davantage les hypothèses sur le système de production et son mode d'exploitation. La connaissance des Φ_i est néanmoins

nécessaire pour pouvoir estimer pendant la phase de programmation si un programme de production est faisable. L'intérêt d'une évaluation rapide de Φ apparaît ici clairement.

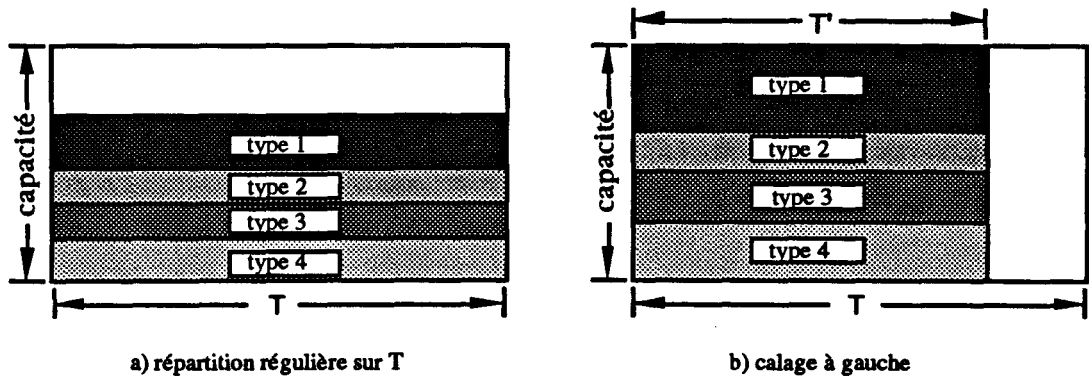


Fig. II-3 ratios proportionnels aux besoins

La première approche est utilisable si les pannes de machines ne doivent pas être prises en compte et si la répartition régulière sur toute la période peut être justifiée, par exemple par l'existence d'une demande constante. L'avantage de la deuxième approche est de constituer progressivement une marge qui est susceptible d'absorber des pannes éventuelles. En revanche, comme la production est effectuée au plus tôt, il peut y avoir une accumulation de produits finis dans le système. Pour des raisons de limitation d'espace de stockage par exemple, une telle accumulation peut être indésirable.

II.2.3 Conduite optimale de flux en présence de pannes

Une méthode de détermination des taux de production, si les niveaux des stocks doivent être contrôlés, a été développée dans [KIM 83]. Elle a suscité un grand intérêt et a été intégrée dans de nombreux travaux, par exemple dans une approche de conduite d'atelier utilisant une modélisation par Réseaux de Petri [LON 93]. Des extensions de la méthode concernent notamment la prise en compte de gammes flexibles et une différenciation des pannes selon leurs fréquences d'apparition [XIE 89][XIE 91].

La méthode est basée sur un calcul de commande optimale des taux de production des différents types de produits et tient explicitement compte des pannes de machines.

Nous illustrons cette méthode dans la suite en utilisant les notations suivantes qui complètent celles introduites précédemment:

- M nombre de machines, indicées par $m = 1..M$
- $\alpha(t)$ vecteur d'état des machines, $\alpha(t) = [\alpha_1(t), \alpha_2(t), \dots, \alpha_M(t)]^T$
avec $\alpha_m(t) = 0$, machine m en panne
 $\alpha_m(t) = 1$, machine m en marche*
 $\alpha_0 = \alpha(t_0)$, désigne l'état initial des machines
- E ensemble des états atteignables, $\text{Card}(E) = 2^M$
- $d(t)$ vecteur de taux de demande, $d(t) = [d_1(t), d_2(t), \dots, d_I(t)]^T$
- $D(t)$ vecteur de demande cumulée, $D(t) = [D_1(t), D_2(t), \dots, D_I(t)]^T$
$$D_i(t) = \int_0^t d_i(s) ds \quad i = 1..I$$
- $\Phi(t)$ vecteur de taux de production, $\Phi(t) = [\Phi_1(t), \Phi_2(t), \dots, \Phi_I(t)]^T$
- $U(t)$ vecteur de production cumulée, $U(t) = [U_1(t), U_2(t), \dots, U_I(t)]^T$
$$U_i(t) = \int_0^t \Phi_i(s) ds \quad i = 1..I$$
- $x(t)$ vecteur d'état du stock de produits, $x(t) = [x_1(t), x_2(t), \dots, x_I(t)]^T$
 x_0 désigne l'état initial du stock et $x(t) - x(0) = U(t) - D(t) \quad i = 1..I$
- K_i nombre d'opérations sur un produit de type i
- y_{im}^k taux de visite des produits de type i sur la machine m pour l'opération k
- τ_{im}^k temps de l'opération k d'un produit de type i sur la machine m

Les hypothèses autres que celles sur les échelles temporelles énoncées dans la section précédente, concernent d'un côté le processus et de l'autre côté la demande à laquelle le système est soumis. Pour le processus, la loi de pannes des machines est supposée exponentielle, ce qui implique que la probabilité qu'une machine change d'état ne dépend que de l'état dans lequel elle se trouve ainsi que d'une probabilité de transition:

$$\text{Prob}[\alpha_m(t + \delta t) = b \mid \alpha_m(t) = a] = \lambda_{ab} \delta t + o(\delta t) \quad (\text{II.5})$$

* dans [KIM 83] le modèle est étendu au cas des ressources multiples. Les composants de $\alpha(t)$ représentent alors le nombre de ressources de type m en état de marche.

L'hypothèse markovienne implique en particulier qu'il ne peut y avoir qu'un seul changement d'état pendant un temps δt infiniment petit. Dans le cas où une machine m passe de l'état $\alpha_m(t) = 1$ à l'état $\alpha_m(t) = 0$, le taux moyen de transition λ correspond à l'inverse du MTTF. De même, une réparation avec un passage de $\alpha_m(t) = 0$ à $\alpha_m(t) = 1$ est régie par un taux qui est l'inverse du MTTR* .

Pour que la modélisation par flux respecte les contraintes technologiques, les trois contraintes suivantes sont imposées:

$$y_{im}^k \geq 0 \quad i = 1..I, m = 1..M, k = 1..K_i \quad (\text{II.6})$$

$$\sum_m y_{im}^k = \Phi_i(t) \quad i = 1..I, k = 1..K_i \quad (\text{II.7})$$

$$\sum_{ik} y_{im}^k \tau_{im}^k \leq \alpha_m(t) \quad m = 1..M \quad (\text{II.8})$$

La contrainte (II.6) exprime le fait que le flux de pièces ne peut pas devenir négatif. La contrainte (II.7) impose la conservation du flux pour chaque type de produit et pour chaque opération d'une gamme. Etant donné que la sommation est faite sur toutes les machines, cette formulation inclut le cas où plusieurs machines sont susceptibles d'exécuter la $k^{\text{ème}}$ opération d'un produit de type i . La contrainte (II.8) exprime la limitation du flux total par la capacité des machines. En cas de panne, cette contrainte assure l'annulation de tous les flux de produits traversant la machine concernée.

Une commande $\Phi(t)$ est alors dite admissible s'il existe des y_{im}^k , $i = 1..I$, $m = 1..M$, $k = 1..K_i$, qui respectent les contraintes (II.6), (II.7) et (II.8). Le domaine des commandes admissibles est noté $\Omega(t)$. En ce qui concerne la demande, elle peut être une fonction $d(t)$ quelconque dans le cas général. Cependant, pour des raisons de complexité de calcul et aussi par défaut de données plus précises, $d(t)$ reste constante dans la plupart des applications de la méthode et ses composantes sont donc données par

$$d_i(t) = \frac{D_i(T)}{T} \quad i = 1..I \quad (\text{II.9})$$

* dans le cas des ressources multiples, le processus de pannes des machines est alors un processus de naissance-mort indépendant pour chaque type de machines.

L'objectif de la méthode est alors de trouver une commande $\Phi(t)$ qui suit au plus près la demande en anticipant les pannes des machines. Pour satisfaire la demande pendant ces pannes, un stock de produits finis est tenu. Le niveau de ce stock, $x_i(t)$, $i = 1..I$, est contrôlé par deux types de pénalisations, d'un côté pour une rupture de stock, de l'autre côté pour une accumulation de produits. Les hypothèses sur cette fonction de pénalisation sont les suivantes

$$g(x) = \sum_i g_i(x_i) \quad (\text{II.10})$$

$$\lim_{|x_i| \rightarrow \infty} g_i(x_i) = \infty \quad i = 1..I \quad (\text{II.11})$$

$$g_i(0) = 0 \quad i = 1..I \quad (\text{II.12})$$

Une fonction de pénalisation simple qui remplit les conditions (II.10), (II.11) et (II.12) est par exemple linéaire par morceaux pour chaque composante i :

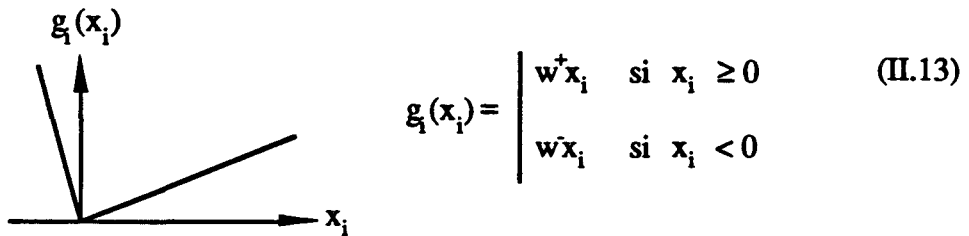


Fig. II-4 fonction de pénalisation linéaire par morceaux

L'optimisation peut être effectuée sur un horizon fini ou infini. Dans le cas où l'horizon est fini, le problème consiste alors à déterminer la commande $\Phi(t)$ qui minimise l'espérance mathématique de la valeur cumulée de la fonction de pénalisation:

$$\text{minimiser } J(x, \alpha, t_0 = 0) = E \left\{ \int_{t_0}^T g(x(s)) ds \mid x(0) = x_0 \text{ et } \alpha(0) = \alpha_0 \right\} \quad (\text{II.14})$$

$$\text{avec } \frac{dx(t)}{dt} = \Phi(t) - d(t) \quad \Phi(t) \in \Omega(t)$$

$$\text{Prob}[\alpha_m(t + \delta t) = b \mid \alpha_m(t) = a] = \lambda_{ab} \delta t + o(\delta t) \quad (a, b) \in E \times E$$

En appliquant le principe de Bellmann, qui dit que toute partie d'un chemin optimal est elle-même optimale, il est possible d'isoler des parties élémentaires de longueur δt . Pour des δt suffisamment petits, l'intégrale peut être approchée par le produit de l'intégrand par δt . De même, l'accroissement de la fonction J peut être approché par un développement au premier ordre. Moyennant des simplifications, la commande optimale est solution de l'expression suivante:

$$\Phi^* = \arg \min_{\Phi \in \Omega(t)} \left[\left(\frac{\partial J(x, \alpha, t)}{\partial x} \right)^T (\Phi(t) - d) \right] \quad (\text{II.15})$$

Etant donné la difficulté de la résolution de cette équation partielle de Bellmann dans le cas général, la fonction J est souvent approximée par une fonction quadratique. La commande obtenue avec une telle approximation n'est donc pas optimale, mais donne de résultats satisfaisants en pratique [LON 93]. Les différentes phases qui peuvent exister dans la poursuite de la consigne, indépendamment pour chaque type de produit, sont représentées dans la figure II-5.

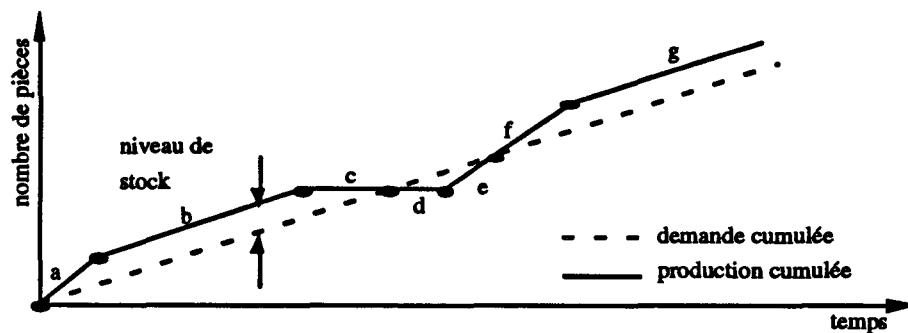


Fig. II-5 Poursuite d'une demande constante

Pendant les phases (a), (e) et (f) le taux de production est supérieur à la demande, soit pour rattraper un retard, soit pour constituer un stock de sécurité. Les phases (c) et (d) correspondent à des pannes pendant lesquelles la production du type de produit représenté est arrêtée avec une rupture de stock en (d) qui persiste jusqu'à la fin de (e). En (b) et (g) la production correspond exactement à la demande et le stock de sécurité est maintenu à un niveau constant.

La représentation de la figure II-5 est toutefois simplifiée, car le taux de production maximal dépend en général des états de toutes les ressources ainsi que des niveaux de stocks de tous les produits. Cette dépendance provient des ressources partagées entre plusieurs gammes de produits. Ce partage de ressources, qui est explicité et chiffré par les contraintes (II.7) et (II.8), peut se traduire par des taux de production et donc des pentes en (a), (e) et (f) variables.

Puisque l'objectif de la méthode est de contrôler les niveaux des stocks de produits finis, elle se prête bien pour une production en *flux tendu* où tous les produits sont fabriqués en même temps. Cela suppose néanmoins que le nombre de types de produits différents soit limité car une fabrication simultanée de tous les types de produits implique que les magasins d'outils des machines d'usinage doivent contenir tous les outils nécessaires. De même, des palettes avec des supports de bridage pour tous les types de produits doivent être disponibles sur le convoyeur.

Si le nombre de types de produits différents est trop grand, des contraintes de capacité, aussi bien au niveau des magasins d'outils des machines à commande numérique* qu'au niveau du nombre de types de palettes, peuvent interdire la fabrication simultanée de tous les types de produits. Dans ce cas, la méthode, telle qu'elle a été présentée n'est plus applicable. Un autre problème de l'approche réside dans le fait que les valeurs numériques des ratios de production issues du calcul d'optimisation peuvent être a priori quelconques. Les ratios de production ne peuvent pas être obtenus par la relation (II.3) parce que la production n'est pas forcément à son maximum. Ils sont cependant définis par :

$$r_i = \frac{\Phi_i}{\sum_i \Phi_i} \quad i = 1..I \quad (\text{II.16})$$

Les seules relations que les r_i vérifient sont:

$$r_i \geq 0 \quad i = 1..I \quad (\text{II.17})$$

$$\sum_i r_i = 1$$

* dans le cas d'un assemblage de composants électroniques cette contrainte correspond au nombre limité d'emplacements pour les alimentations en composants.

Les valeurs numériques de ces ratios de production jouent cependant un rôle important pendant l'élaboration d'un ordonnancement cyclique. Nous allons voir par la suite que d'autres contraintes, liées à des aspects de faisabilité et de performances, doivent être prises en compte pendant la détermination des ratios.

II.2.4 Les approches de regroupement

Afin de tenir compte des limitations de capacité qui peuvent exister au niveau des capacités des magasins d'outils ou des palettes, différentes approches de regroupement de types de produits ont été développées. Selon [STE 88], la problématique de regroupement comprend les cinq points suivants:

- (1) la sélection des produits à fabriquer simultanément
- (2) le regroupement des machines en groupes de configurations identiques
- (3) la détermination des ratios de production des différents produits
- (4) l'allocation des palettes et supports de bridage
- (5) l'allocation des outils aux magasins des machines

Plusieurs types d'approche de regroupement peuvent être distingués. Parmi eux il y a les approches de regroupement en familles, appelées aussi *batching* [HWA 86] [MAZ 89]. Elles consistent à regrouper en familles, les différents types de produits à fabriquer simultanément. Tous les produits d'une même famille sont alors fabriqués sans interruption jusqu'à ce que tous les besoins de fabrication pour tous les produits soient satisfaits. Une phase de reconfiguration est alors engagée qui consiste à changer les outils et les supports de bridage - la broche d'un tour par exemple - pour préparer les machines à la production de la famille suivante.

Dans [STE 88] sept différentes approches de regroupement en familles sont comparées. L'objectif commun de toutes ces approches est la minimisation du nombre de familles différentes, et ainsi du nombre de reconfigurations. Cet objectif est justifié par l'hypothèse qu'un atelier flexible constitue souvent le goulot d'étranglement d'une fabrication et qu'il est par conséquent utilisé au voisinage de sa capacité maximale. Un trop grand nombre de reconfigurations pourrait alors le rendre incapable d'absorber la

charge programmée [MAZ 89]. L'équilibrage des charges des machines est cependant considéré comme un critère secondaire dans toutes les approches de regroupement comparées. Il est alors possible que cette charge soit mal équilibrée à l'intérieur d'une famille. Selon [STE 92], la charge peut être répartie de plusieurs façons, qui correspondent à des objectifs différents: dans la représentation de la figure II-6 les besoins de fabrication sont répartis régulièrement de sorte que tous les types de produits commencent et terminent simultanément leur production.

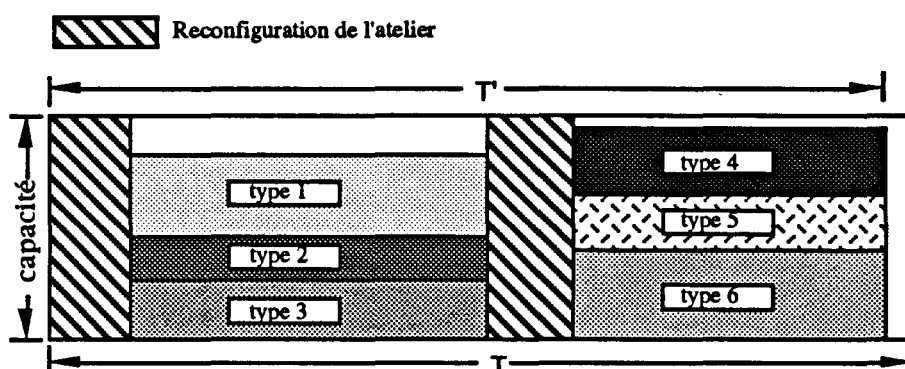


Fig. II-6 Regroupement en familles (batching)

Une telle approche est adaptée si la demande est *dépendante*, pour un assemblage à un poste suivant par exemple. Un autre objectif consiste à caler la production le plus possible à gauche afin d'utiliser les machines au maximum pendant un premier temps. Avec cette approche une marge est progressivement dégagée, au moins pour certains produits, comme indiqué sur la figure II-7.

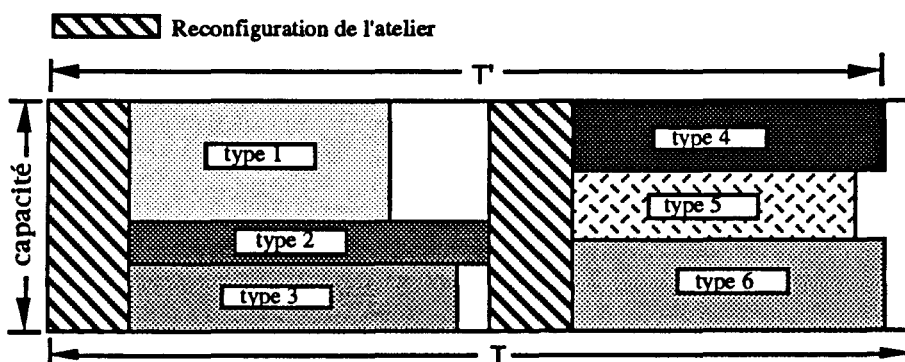


Fig. II-7 Regroupement en familles avec calage à gauche

Les inconvénients des approches de regroupement en familles se situent à deux niveaux. Premièrement, la charge de travail pour le personnel qui effectue la configuration est très ponctuelle. Pendant une reconfiguration toute la production est arrêtée bien que les travaux de configuration s'effectuent rarement sur toutes les machines en même temps. Le deuxième inconvénient provient du fait que l'unique objectif poursuivi dans ces approches est la minimisation du nombre de reconfigurations. L'équilibrage de la charge des machines est alors négligé et une mauvaise utilisation de l'atelier peut en résulter.

Dans [STE 91] une nouvelle approche de sélection de types de pièces et de détermination des ratios est proposée. Cette approche, qualifiée de *regroupement dynamique*, tient compte des contraintes sur les capacités de magasins d'outils et tente en même temps d'équilibrer la charge des machines afin de maximiser la production. L'idée de base est de permettre des reconfigurations partielles de l'atelier. Les machines qui ne sont pas concernées par une reconfiguration peuvent continuer à produire pendant ce temps. Une reconfiguration avec une nouvelle détermination des types à produire et des ratios de production associés est effectuée si:

- les besoins de fabrications pour un type de produits sont satisfaits
- une machine tombe en panne
- une machine est remise en service
- une commande urgente arrive
- une maintenance préventive est effectuée

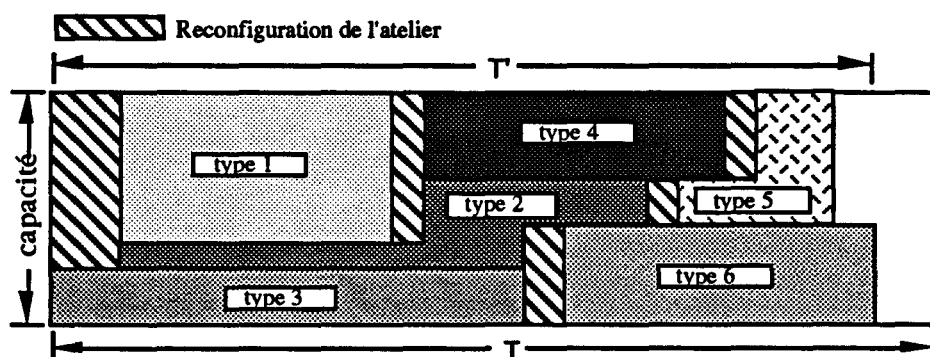


Fig. II-8 Regroupement dynamique [STE 91]

Contrairement aux approches de regroupement en familles, l'optimisation est effectuée chaque fois qu'un de ces événements a lieu. Cette optimisation ne concerne alors que la période de temps allant jusqu'au prochain événement et une optimalité globale ne peut donc pas être espérée. Nous allons illustrer l'approche de regroupement en familles en utilisant les notations suivantes:

M	nombre de machines, $m = 1..M$
K	nombre de types de machines, $k = 1..K$
m_k	nombre de machines de type k
C	nombre d'outils, $c = 1..C$
I	nombre de types de produits, $i = 1..I$
p_{ik}	temps opératoire d'un produit de type i sur une machine de type k
\bar{p}_{ik}	temps opératoire d'un produit de type i réparti régulièrement sur toutes les m_k machines de type k : $\bar{p}_{ik} = p_{ik} / m_k$
e_{ck}	nombre d'emplacements nécessaires pour un outil de type c sur une machine de type k
v_{ick}	$= 1$, si l'usinage d'un produit de type i sur une machine de type k requiert un outil de type c $= 0$, sinon
t_k	nombre d'emplacements d'un magasin d'outils d'une machine de type k
C_{k1}	pondération pour une sur-charge des machines de type k
C_{k2}	pondération pour une sous-charge des machines de type k
f_i	entier proportionnel au ratio de production pour les produits de type i
F_i	borne supérieure pour f_i
W_k	charge de référence pour une machine de type k (en unité de temps)
X_{k1}	sur-charge d'une machine de type k (en unité de temps)
X_{k2}	sous-charge d'une machine de type k (en unité de temps)
y_{ck}	$= 1$, si l'outil c est chargé sur une machine de type k $= 0$, sinon

Le problème de sélection des types de produit et de leur ratios associés se présente alors sous la forme d'un problème de programmation en variables entières:

$$\text{minimiser } \sum_k C_{k1} X_{k1} + \sum_k C_{k2} X_{k2} \quad (\text{II.18})$$

$$\text{sous les contraintes } \sum_i \tilde{p}_{ik} f_i - X_{k1} + X_{k2} = W_k \quad k = 1..K \quad (1)$$

$$v_{ick} f_i \leq F_i y_{ck} \quad i = 1..I, k = 1..K, c = 1..C \quad (2)$$

$$\sum_c e_{ck} y_{ck} \leq t_k \quad k = 1..K \quad (3)$$

$$f_i \leq F_i \quad i = 1..I \quad (4)$$

$$f_i \geq 0 \quad i = 1..I \quad (5)$$

$$X_{k1}, X_{k2} \geq 0 \quad k = 1..K \quad (6)$$

Parmi les paramètres utilisés, les W_k jouent des rôles particuliers. Leur unité est le temps et ils indiquent une charge de référence pour chaque type de ressource. La minimisation de (II.18) consiste alors à déterminer des f_i de sorte que les charges de toutes les ressources soient équilibrées. Avec $W_k = W$, $k = 1..K$, tous les types de machines ont la même charge de référence, ce qui correspond à un équilibrage entre les différents types de machines. Si nous supposons que les variables d'écart X_{k1} et X_{k2} sont petites, les W_k sont à peu près proportionnels aux f_i qui sont les variables de décision à déterminer. Ces variables f_i correspondent au nombre de fois qu'un produit de type i apparaît pendant la période de durée W_k et sont liées au ratios de production par la relation

$$r_i = \frac{f_i}{\sum_i f_i} \quad i = 1..I \quad (\text{II.19})$$

Les contraintes de type (2) imposent que si un produit de type i est effectivement sélectionné, alors les machines par lesquelles ces produits passent doivent contenir les outils nécessaires*. Les contraintes de type (3) expriment la limitation du nombre d'emplacements dans les magasins d'outils. Les valeurs possibles pour les f_i sont limitées par les contraintes de type (4) et (5) et la positivité des variables d'écart est imposée par les contraintes (6). Une variante du problème pour tenir compte d'une demande dépendante

* dans [STE 88] la formulation initialement donnée est: $v_{ick} f_i \leq M y_{ck}$ avec un paramètre M qui doit être 'suffisamment' grand. Nous l'avons remplacée ici par une formulation plus précise, car les F_i sont définis.

entre certains types de produits consiste à ajouter des contraintes de proportionnalité du type

$$f_i = \alpha_{ij} f_j \quad i, j \in I' \quad (\text{II.20})$$

Une autre variante de cette approche [STE 92] consiste à remplacer dans les contraintes de type (5) les inégalités larges par des inégalités strictes

$$f_i > 0 \quad i \in I'' \quad (\text{II.21})$$

Cette formulation impose la production simultanée de certains types de produits. Elle peut être appropriée dans des situations où pour des raisons diverses - des priorités par exemple - la sélection des types de produits doit être contrôlée. Remarquons que des contraintes du type (II.21) peuvent être tellement fortes, que le problème (II.18) ne possède plus aucune solution.

Pour la résolution du problème il faut commencer par fixer les paramètres C_{k1} , C_{k2} et W_k . En ce qui concerne les deux premiers, seul leur rapport est significatif et peut être pris égal à 1 dans un premier temps. Pour les W_k , il est suggéré d'essayer des valeurs différentes. Plus les W_k sont grands, plus il y a de combinaisons de f_i qui vérifient les contraintes de type (1). Comme l'espace des solutions devient alors plus large, le temps de calcul augmente également. Pour limiter le temps de calcul d'une résolution en nombres entiers, une relaxation des contraintes de $f_i \in \mathbb{N}$ à $f_i \in \mathbb{R}$, avec un arrondi des valeurs obtenues est proposée dans [STE 91]. Cette approximation donne de résultats acceptables par rapport à une résolution directe dans le domaine des entiers.

Les valeurs maximales des f_i sont toutefois limitées par les paramètres F_i dans les contraintes de type (4). Dans [STE 91] et [STE 92] ces contraintes sont utilisées pour exprimer des limitations du nombre de palettes disponibles. Cette interprétation des contraintes de type (4) est liée au fait que les auteurs envisagent une réalisation des ratios avec un nombre de palettes proportionnel aux f_i . Nous allons montrer dans le chapitre III que cette corrélation entre les ratios et le nombre de palettes n'est pas justifiée, mais que les contraintes de type (4) peuvent cependant avoir une autre utilité.

Une comparaison de l'approche par regroupement dynamique avec l'approche de conduite optimale de flux présentée précédemment est difficile car les objectifs des deux approches sont différents. La première cherche à maximiser la production, ou au moins à satisfaire un besoin de production le plus vite possible sans tenir compte des niveaux de stocks. L'objectif de la deuxième approche est de minimiser une fonction de coût intégrant des coûts de stock et de rupture. Malgré cette différence dans les objectifs poursuivis, nous retrouvons des considérations caractéristiques d'une comparaison entre une commande robuste et une commande optimale en automatique.

L'approche par regroupement dynamique ne nécessite pas d'hypothèses précises sur les occurrences et durées des pannes et est capable de s'accommoder de perturbations très différentes. Elle consiste à effectuer une optimisation sur un horizon court parce que les incertitudes interdisent de toute façon d'effectuer des prévisions détaillées sur un horizon long. En revanche, l'approche de conduite optimale de flux requiert des hypothèses explicites sur la loi des pannes et des réparations. Elle intègre ces informations dans le calcul de la commande en anticipant les pannes de machines par le contrôle du niveau d'un stock de sécurité.

Il apparaît alors que l'approche par regroupement dynamique permet une formulation réaliste du problème de la planification fine, car elle est capable de tenir compte d'une demande dépendante et surtout des contraintes sur les capacités des magasins d'outils. Un autre avantage de l'approche par regroupement dynamique est de pouvoir influencer les valeurs numériques des ratios obtenus, grâce aux contraintes de type (4). En revanche, la formulation telle qu'elle a été présentée en (II.18) ne permet pas de tenir compte de flexibilités autres que celles relatives à l'ordre des opérations, car les temps opératoires des produits sur les différentes machines (p_{ik}) doivent être explicités.

L'étude fine que nous allons présenter nécessite une modélisation plus détaillée qui permet d'obtenir des résultats plus précis. Nous allons donc introduire des approches de modélisation et d'évaluation de performances qui intègrent le caractère discret, négligé jusqu'à maintenant.

II.3 Modèles et techniques pour l'évaluation des performances des systèmes à événements discrets

II.3.1 Prise en compte des événements discrets

La planification fine que nous venons de présenter dans la section précédente s'est limitée à une approche par flux continu pour éviter la complexité inhérente aux systèmes à événements discrets. Les résultats de cette étape sont les valeurs numériques des ratios de production qui peuvent être spécifiés directement, ou bien indirectement à travers des taux de production par des relations de type (II.16). A ce stade, l'introduction d'un modèle tenant compte du caractère discret peut s'effectuer selon plusieurs perspectives:

Premièrement, il peut exister un besoin d'évaluation des performances de l'atelier pour la *vérification d'une planification fine*, préalablement conçue avec les approches approximatives précédemment présentées. Puisque les approches de planification fine utilisent des hypothèses simplificatrices, notamment une approche par flux, il n'est pas du tout certain que le système réel se comporte effectivement comme souhaité. Des contraintes liées au caractère discret et dynamique des systèmes considérés peuvent effectivement dégrader les performances et ainsi remettre en cause l'existence d'une planification fine. Afin de faciliter la prise de décision qui est en rapport avec une planification fine, la confirmation de dates de livraison ou la mise à disposition de pièces brutes pour une certaine date par exemple, il peut s'avérer nécessaire d'analyser le comportement du système sur un modèle intégrant des aspects négligés auparavant. Cela ne signifie pas pour autant que le modèle utilisé soit rendu complètement déterministe.

Deuxièmement, en partant des résultats de la planification fine, l'étape suivante peut consister à raffiner le modèle dans la perspective de *l'élaboration d'une commande déterministe*. Cette commande sera élaborée en tenant compte d'un ou plusieurs objectifs spécifiés, la maximisation de la production par exemple. Dans le cas d'un système sans ou avec peu de perturbations, il s'agit de déterminer un ordonnancement au sein duquel les conflits sont résolus de manière prévisionnelle. En revanche, si le système subit de fortes perturbations, l'élaboration d'un ordonnancement prévisionnel n'est pas adaptée car la probabilité de pouvoir effectivement l'appliquer est trop faible. Une autre solution consiste alors à envisager une *conduite réactive* dans laquelle les conflits sont résolus avec des règles de décisions.

Les différences entre ces perspectives se retrouvent au niveau des modèles associés. Pour une évaluation des performances ayant pour but de vérifier une planification fine, une résolution probabiliste des conflits peut être adaptée et constituer une approximation suffisante de la réalité*. En revanche, l'élaboration d'une commande consiste à résoudre définitivement les indéterminismes et l'issue d'une telle approche sera par conséquent un modèle déterministe de type graphe d'événements par exemple. En ce qui concerne la conduite réactive, elle doit être aussi réalisée dans le cadre d'une approche de modélisation adaptée. Le modèle obtenu conservera cependant des indéterminismes.

II.3.2 La simulation

La simulation est l'outil d'évaluation de performances par excellence pour les systèmes à événements discrets complexes qui ne peuvent être étudiés avec des méthodes analytiques. Simuler consiste à reproduire l'évolution dans le temps d'un système, ou plus précisément d'un modèle d'un système. La simulation est une technique d'analyse qui s'applique a priori à tout modèle d'un système à événements discrets. Il existe donc des simulateurs basés sur des Réseaux de Petri (GSPN), des Réseaux de files d'attente (QNAP) ou basés sur des approches de modélisation spécifiques (SLAM, SIMAN, Witness, ...) [PRO 87][CER 88]. Les simulateurs basés sur des Réseaux de Petri ou des Réseaux de files d'attente permettent souvent de combiner une phase d'évaluation analytique avec une évaluation par simulation.

L'approche par simulation, surtout avec une modélisation spécifique à la problématique traitée, permet l'analyse très précise d'un système réel. Ceci provient du fait que l'évaluation d'un modèle par simulation n'introduit pas d'hypothèses restrictives pour la modélisation. La facilité de modéliser avec finesse pratiquement tout système constitue le grand atout de la simulation et explique son emploi fréquent en pratique. Ainsi, des stratégies de conduite réactive avec des règles qui dépendent de l'état du système, par exemple des longueurs de files d'attente ou des temps opératoires rémanents des machines, peuvent seulement être évaluées avec la simulation. De même, les régimes transitoires et les problèmes liés à des non-linéarités, comme des lois de pannes dépendantes de l'intensité d'utilisation des machines, ou des blocages dus à des limitations de

* la vérification se rapporte alors au modèle utilisé. Une vérification dans le sens 'faisabilité réelle' est difficilement justifiable sur un modèle stochastique.

capacités de stockage, pour n'en citer que quelques uns, peuvent être modélisés et évalués efficacement par simulation.

Cette facilité de mise en œuvre risque néanmoins de cacher des problèmes liés à la simulation et à l'interprétation des résultats obtenus. Quelques uns de ces problèmes peuvent être facilement évités parce qu'ils résultent d'une mauvaise utilisation de la simulation. Un exemple simple d'illustration est relatif à la traduction directe des fractions utilisées au niveau flux en probabilités pour les décisions au niveau discret. De même, la simulation est fréquemment utilisée pour une *validation* de propriétés qualitatives dans un contexte stochastique, l'absence de blocage par exemple, sans que la complétude de la simulation, c'est-à-dire le passage par tous les états possibles du système, puisse être garantie*.

D'autres types de problèmes, inhérents à l'approche par simulation, proviennent de la prise en compte d'événements rares. Si ces événements rares relèvent d'une distribution probabiliste, il faut simuler un grand nombre de ces événements pour approcher correctement la fonction de distribution. Entre deux événements rares il y a cependant un grand nombre d'autres événements de sorte que le temps de simulation puisse devenir très long. La problématique des événements rares se rencontre dans l'analyse de systèmes de production assujettis à des pannes de machines. Avec des pannes rares et des durées de pannes très variables, la simulation peut parfois concerner quelques centaines d'heures de fonctionnement avant de permettre l'obtention de résultats stables. Dans une telle situation il faut se demander si l'horizon de simulation a été convenablement choisi et s'il ne vaut pas mieux se placer dans un régime stable, c'est-à-dire dans une période de type (3) selon les hypothèses de Gershwin.

Le problème du choix de la durée de simulation ne se pose pas seulement en présence d'événements rares, mais il constitue un problème général de l'approche. En effet, la simulation est une technique purement expérimentale qui fait correspondre à un jeu de paramètres et de données d'entrée des résultats sous forme de valeurs numériques. Il est donc impossible de spécifier des performances, un débit de pièces par exemple, et d'en

* il est facile de construire un exemple simple d'un système discret qui se bloque seulement après un temps de simulation aussi grand qu'on le veut. Le Réseau de Petri de la figure II-12 avec une résolution de conflit équiprobable, se bloque avec une probabilité qui dépend du nombre de marques initiales.

déduire par une approche générative les valeurs des paramètres nécessaires pour l'obtention de ces performances.

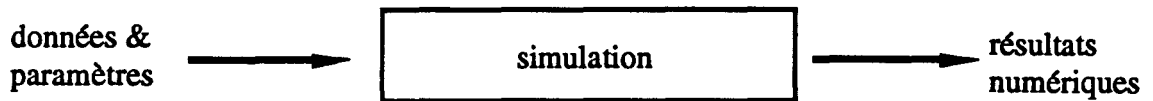


Fig. II-9 La simulation comme approche purement évaluative

En présence d'un grand nombre de paramètres, l'explosion combinatoire du nombre de jeux d'essais qu'il faudrait tester pour déterminer les valeurs optimales des paramètres, rend de fait la simulation impraticable. Des approches d'analyse de perturbation, où l'influence d'une variation des paramètres est estimée, permettent dans certains cas de réduire le nombre de simulations [HO 83][HO 87][SUR 89]. Dans le cas général, il faut néanmoins avoir recours à des approches analytiques pour déterminer les paramètres par un calcul génératif ou au moins réduire le temps de calcul de telle sorte qu'un grand nombre de configurations puisse être évalué pendant un temps de calcul acceptable.

II.3.3 Les approches analytiques

La caractérisation des approches analytiques porte aussi sur les deux aspects *modélisation* et *résolution* qui sont évidemment étroitement liés. Les approches de modélisation les plus répandues pour l'analyse des systèmes à événements discrets sont les Réseaux de Files d'Attentes (RFA) et les Réseaux de Petri (RdP). Une comparaison détaillée de ces approches de modélisation peut être trouvée dans [FDI 89] par exemple et nous nous limitons ici à une description plus restreinte, dans l'optique de la problématique traitée.

L'élément de base de la théorie des files d'attente est la *station* qui est composée d'un ou de plusieurs *serveurs* et d'une *file d'attente* associée dans laquelle des *clients* attendent la disponibilité d'un serveur. Pour les serveurs il faut spécifier la loi de service et pour la file d'attente le mode de gestion utilisé (FIFO, LIFO, ...). Plusieurs stations peuvent former un *réseau* à l'intérieur duquel des clients circulent. A chaque station le

flux de clients sortant est routé vers une autre station ou partagé pour aller vers des stations différentes. Si le réseau est ouvert, des arrivées de clients de l'extérieur et des départs vers l'extérieur sont pris en compte, sinon le nombre de clients à l'intérieur du réseau reste constant. Pour la modélisation d'ateliers flexibles des réseaux ouverts et fermés ont été utilisés [SOL 79][DAL 84]. Dans les réseaux ouverts, les clients correspondent aux produits qui passent par l'atelier, tandis que dans les réseaux fermés les clients modélisent les palettes, supports des produits. Cette deuxième approche est généralement préférée car le nombre de clients dans le système est alors constant et l'hypothèse de la finitude des capacités des files d'attente n'est pas nécessaire.

Initialement la modélisation était limitée à une seule classe de clients. Une extension à des clients appartenant à des classes différentes a été présentée dans le théorème BCMP [BAS 75], mais requiert une gestion des files d'attente en mode processeur partagé - ce qui renvoie d'une certaine manière au calcul de flux. De même, des primitives de synchronisation et de limitation de capacité ont été ajoutées aux primitives de modélisation initiales, mais la résolution analytique se limite alors à des cas particuliers. Dans le cas général, ces primitives de modélisation entraînent alors une résolution par simulation.

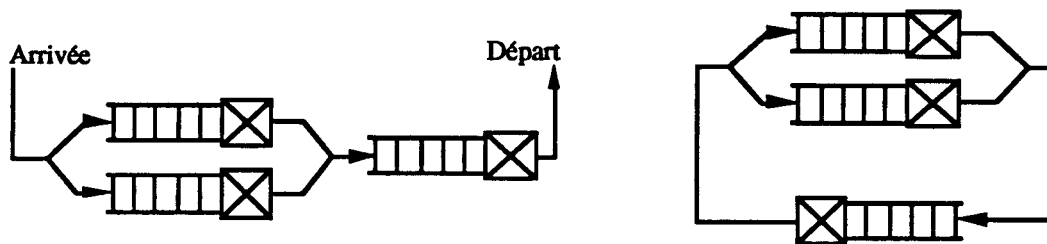


Fig. II-10 Exemple de réseaux de files d'attente ouvert et fermé

La technique de résolution analytique la plus répandue associée à ce modèle est basée sur l'hypothèse que le processus de changement d'état, c'est-à-dire le passage d'un client d'une station à une autre, s'identifie à un processus de Markov. Il faut alors supposer que les temps de service des serveurs suivent une distribution exponentielle et que la résolution des conflits de routage est effectuée de manière probabiliste. L'hypothèse de la finitude des capacités des files d'attente ne concerne que les réseaux ouverts car le nombre de clients dans une file d'attente dans un réseau fermé est de toute façon borné par le nombre total de clients dans le système. Dans le cas d'une gestion des files en mode FIFO, un seul type de client est admis. Une approximation souvent utilisée pour

tenir compte de clients de types différents consiste à introduire des temps de service moyens calculés comme étant la moyenne des temps de service des différents types de clients pondérée par les taux de visite respectifs.

Dans les réseaux fermés qui vérifient les hypothèses énoncées ci-dessus, la probabilité que le système se trouve dans un certain état $n = (n_1, n_2, \dots, n_M)$ s'exprime alors sous la forme d'un produit

$$P(n_1, n_2, \dots, n_M) = K P(n_1) P(n_2) \dots P(n_M) \quad (\text{II.22})$$

Les facteurs $P(n_i)$ s'évaluent directement en fonction des données du problème et il existe également des algorithmes rapides pour la détermination de la constante de normalisation K . Cette possibilité de résolution immédiate constitue le véritable intérêt de l'approche par réseaux de files d'attente. A partir des probabilités des différents états atteignables se déduisent alors facilement des mesures de performances telles que les débits des stations, les longueurs moyennes des files, les temps moyens d'attente et les taux d'utilisation moyens des serveurs.

Des extensions de ce résultat fondamental ont été développées selon plusieurs axes et concernent notamment l'intégration de contraintes de synchronisation [BAC 89], la prise en compte de lois de services autres qu'exponentielles [MAR 79] ou la limitation des capacités des files [BOU 92]*. Une approche d'analyse différente, toujours basée sur une modélisation par files d'attente a été proposée dans [DAL 84]. Cette approche appelée *analyse opérationnelle* utilise des hypothèses beaucoup plus faibles, en particulier sur les temps de service, et s'appuie sur les grandeurs observables à partir desquelles un certain nombre de mesures de performances peut être déduit.

L'évaluation des performances d'un atelier flexible basée sur une modélisation par files d'attentes est envisagée dans [STE 92]. Les avantages sont une prise en compte facile des ratios de production et une résolution immédiate par la forme produit. Ainsi, le comportement de l'atelier soumis à une certaine charge de travail peut être évalué afin de s'assurer par exemple qu'un programme de production soit effectivement faisable.

* voir [HSU 93] pour une étude bibliographique abondante et une classification des résultats analytiques en fonction des hypothèses utilisées.

Les inconvénients d'une évaluation par réseaux de files d'attente sont cependant nombreux et se situent à plusieurs niveaux que nous évoquons ci-après:

◆ En premier lieu, les hypothèses nécessaires pour une résolution efficace utilisant la forme produit sont très fortes. Une loi de probabilité exponentielle avec un coefficient de variation égal à un est difficilement justifiable pour la modélisation d'un temps d'usinage, surtout dans le contexte des hypothèses énoncées dans II.2.1 et selon lesquelles un régime stable et non-perturbé existe entre deux pannes de machines. De même, la traduction des ratios de production et des ratios de routage à l'intérieur du système en probabilités, implique que seuls les premiers moments des distributions réelles sont pris en compte correctement. Les approximations utilisées accentuent les variations des distributions, c'est-à-dire surestiment les écarts-types réels, ce qui entraîne généralement des résultats pessimistes.

◆ Un autre inconvénient provient du fait qu'un raffinement ultérieur du modèle n'est pas possible. En termes de perspectives précédemment évoquées et selon lesquelles une évaluation de l'atelier avec une prise en compte du caractère discret peut être effectuée, une utilisation de l'approche par files d'attente se limiterait donc à la phase de validation d'un programme de production. Comme l'ambition du projet CASPAIM est aussi l'élaboration de la commande, il est souhaitable de pouvoir raffiner progressivement un modèle d'évaluation préliminaire, afin de le rendre ultérieurement complètement déterministe. Pour traiter les problèmes d'ordonnancement qui suivent naturellement la phase de programmation, il faudrait alors ne pas changer seulement de modèle, mais totalement d'approche de modélisation.

◆ Un troisième inconvénient est lié au précédent et concerne le niveau conceptuel de modélisation. Le passage de CASPAIM I à CASPAIM II était motivé par une prise en compte équitable des ressources et des produits. Les primitives de modélisation offertes par l'approche réseaux de files d'attentes privilégient cependant l'aspect ressource et l'architecture physique au détriment de l'aspect produit. La modélisation centrée sur des serveurs avec des files d'attentes associées implique de plus que la notion de file d'attente est d'une certaine manière imposée par la modélisation. Dans le chapitre suivant nous allons voir qu'il existe néanmoins des stratégies efficaces pour la gestion d'un atelier flexible et de son en-cours qui ne font appel ni à des notions de FIFO, ni à celle de file d'attente.

Une modélisation basée sur les Réseaux de Petri (RdP), déjà utilisée dans le projet CASPAIM dans des domaines autres que l'analyse quantitative, permet d'éviter la plupart des inconvénients mentionnés ci-dessus. Premièrement, en ce qui concerne l'intégration du facteur temps dans le modèle, des résultats existent pour des RdP avec des temporisations déterministes, exponentielles ou quelconques. Avec des temporisations exponentielles et une résolution des conflits par des probabilités de Bernoulli, la partie évaluation se ramène alors dans le cas général à la résolution de la chaîne de Markov sous-jacente [MOL 82][FLO 85][MAR 89].

De plus, les résultats établis dans le cadre d'une modélisation par réseaux de files d'attente sont transposables sur des modèles RdP. Un réseau de files d'attente se modélise facilement par un RdP stochastique. Dans ce cas les deux approches de modélisations aboutissent au même processus markovien et la résolution par la forme produit s'applique donc aussi au RdP stochastique correspondant.

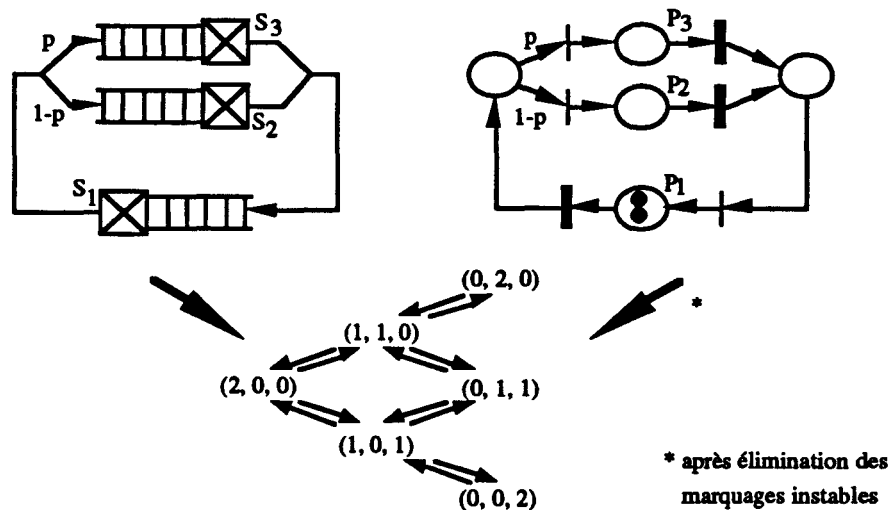


Fig. II-11 Processus de Markov sous-jacent aux RFA et RdP stochastiques

Des temporisations autres qu'exponentielles ont été associées aux RdP. Les temporisations déterministes que nous allons utiliser dans la suite ont été développées par [RAM 80] et il existe également des approches d'analyse opérationnelle basée sur les RdP [CHI 93], similaires à celles utilisées dans le contexte d'une modélisation par réseaux de files d'attente.

L'avantage fondamental d'une modélisation avec des RdP par rapport aux réseaux de files d'attentes usuels apparaît nettement lorsqu'il s'agit de modéliser des systèmes de production flexibles. Il réside principalement dans le fait qu'une primitive de synchronisation est effectivement intégrée dans l'approche. La possibilité de modéliser diverses synchronisations permet d'explicitier des conditions, autre que celles liées aux temporisations, sur l'évolution du système. Des conditions d'accès à des files d'attente à capacités limitées deviennent alors modélisables ainsi que des opérations de type rendez-vous entre produits et machines. Ce dernier point facilite en particulier une modélisation équitable des produits et des ressources.

Au niveau de l'analyse du comportement temporel, la prise en compte de synchronisations a motivé le développement de méthodes d'évaluation spécifiques. Dans [COH 85] [COH 89] et [GAU 90] une approche d'évaluation basée sur une description dans une algèbre avec les deux opérateurs (max , $+$) est présentée. L'opérateur max est parfaitement adapté pour décrire la contrainte de synchronisation tandis que l'opérateur $+$, l'addition classique, permet de tenir compte des retards. Dans cette algèbre l'évolution de systèmes contenant des synchronisations peut alors être décrite par des équations linéaires ce qui permet une représentation sous forme d'équation d'état similaire à celle des systèmes continus. L'approche est cependant limitée à des systèmes sans conflit et s'applique donc principalement à des modèles de type graphes d'événements.

Un autre avantage des RdP est la représentation explicite des marques qui évoluent dans le réseau et qui peuvent avoir différentes interprétations. Dans un même modèle des marques peuvent représenter des objets physiques, comme des produits ou des ressources, ou des objets logiques représentatifs d'une commande. De plus, la dynamique d'un RdP est prise en compte et chaque étape de son évolution peut être représentée graphiquement. Ceci permet en particulier de modéliser et d'analyser des régimes de durées limitées, comme par exemple des régimes transitoires entre des fonctionnements répétitifs.

Remarquons que la prise en compte de synchronisations engendre toute une série de problèmes - en dehors de l'aspect temporisation - concernant notamment la vivacité, la finitude et l'atteignabilité de certains états. Ces propriétés peuvent parfois être caractérisées en fonction de la seule structure d'un réseau, mais une prise en compte explicite du marquage du réseau est nécessaire pour aborder ces problèmes dans le cas général. Con-

trairement aux réseaux de files d'attente qui évoluent de manière autonome grâce aux temporisation exponentielles et résolutions de conflits probabilistes, les RdP ont initialement été développés sans temporisation et sans procédure particulière de résolution de conflits. La définition de la sémantique d'évolution et l'analyse des propriétés occupe par conséquent une plus grande place dans l'approche par RdP.

II.4 Les Réseaux de Petri

II.4.1 Définitions et notations

Les définitions et notations que nous présentons dans la suite sont principalement celles utilisées dans [MUR 89][DIC 93] et [DES 95]. Un Réseau de Petri est constitué de *places* et de *transitions* qui peuvent être reliées par des *arcs* auxquels des *poids* sont associés. Un arc ne relie jamais deux places ou deux transitions. L'ensemble des arcs qui relie des places à des transitions et l'ensemble des arcs qui relie des transitions à des places sont disjoints. La dynamique d'évolution du système est représentée par des *marques* ou *jetons* contenus dans les places.

Un Réseau de Petri est défini par

$$N = \langle P, T, \text{Pré}, \text{Post} \rangle \quad (\text{II.23})$$

avec	$P = \{p_1, p_2, \dots, p_n\}$	ensemble fini de places
	$T = \{t_1, t_2, \dots, t_m\}$	ensemble fini de transitions
	$P \cap T = \emptyset$	
	$\text{Pré}: P \times T \mapsto \mathbb{N}$	fonction de pré-incidence
	$\text{Post}: T \times P \mapsto \mathbb{N}$	fonction de post-incidence

Il existe un arc de la place p_i à la transition t_j si et seulement si $\text{Pré}(p_i, t_j) > 0$. De même, il existe un arc de t_k à p_i si et seulement si $\text{Post}(t_k, p_i) > 0$. Les valeurs de $\text{Pré}(p_i, t_j)$ et de $\text{Post}(t_k, p_i)$ correspondent à des poids associés aux arcs. Si tous les arcs ont un poids égal à un, le réseau est dit *ordinaire*. Les places p telles que $\text{Pré}(p, t) > 0$ (respectivement $\text{Post}(t, p) > 0$) forment l'ensemble d'entrée (respectivement l'ensemble de

sortie) de t qui est noté $\bullet t$ (respectivement $t \bullet$). De même, les transitions t telles que $\text{Post}(t,p) > 0$ (respectivement $\text{Pré}(p,t) > 0$) forment l'ensemble d'entrée (respectivement l'ensemble de sortie) de p qui est noté $\bullet p$ (respectivement $p \bullet$). Le Réseau de Petri est dit *pur* si et seulement si pour toute transition t , $\bullet t \cap t \bullet = \emptyset$.

Les fonctions Post et Pré peuvent être représentées par leurs matrices associées POST et PRE, matrices à n lignes et m colonnes. Ces matrices sont définies par leurs composantes $w_{i,j}^+$ et $w_{i,j}^-$ comme suit

$$w_{i,j}^+ = \begin{cases} \text{Post}(t_j, p_i) & \text{si } p_i \in \bullet t_j \\ 0 & \text{sinon} \end{cases} \quad \text{et} \quad w_{i,j}^- = \begin{cases} \text{Pré}(p_i, t_j) & \text{si } p_i \in \bullet t_j \\ 0 & \text{sinon} \end{cases}$$

Nous désignons par $\text{Post}(t)$ (respectivement $\text{Pré}(t)$) la colonne associée à la transition t dans la matrice POST (respectivement PRE). A partir des matrices POST et PRE, la *matrice d'incidence* du réseau est définie par $W = \text{POST} - \text{PRE}$. La description par W ou par le couple $(\text{POST}, \text{PRE})$ est équivalente si le réseau est pur.

Au 4 - uplet $\langle P, T, \text{Pré}, \text{Post} \rangle$, qui définit la structure statique du réseau, un marquage est associé par la *fonction de marquage* $M: P \mapsto \mathbb{N}$. Le vecteur M (respectivement M_0) représente le marquage instantané (marquage initial) du réseau, $M(p_i)$ (respectivement $M_0(p_i)$) représente le marquage instantané (marquage initial) de la place p_i . Un réseau est alors complètement défini par $R = \langle N, M \rangle$.

Une transition t est dite tirable pour un marquage M si dans chacune de ses places d'entrée le nombre de marques est supérieur ou égal au poids de l'arc reliant cette place à la transition, i.e. $\forall p \in \bullet t, M(p) \geq \text{Pré}(p,t)$. Si ces conditions sont vérifiées, le tir de la transition t engendre le nouveau marquage M' donné par $M' = M + \text{Post}(t) - \text{Pré}(t)$. Le passage d'un marquage M à M' par le tir de t est aussi noté $M[t > M']$.

Une séquence de tirs est notée σ et définie par* $\sigma = \{t_1 \circ t_2 \circ \dots \circ t_k\}$. Un vecteur $\bar{\sigma} \in \mathbb{N}^m$ est associé à chaque séquence de tirs σ et une composante i de ce vecteur représente le nombre de fois qu'une transition t_i apparaît dans σ . Si une séquence est

* l'opérateur \circ représente la concaténation.

tirable, le marquage M' obtenu est alors donné par l'équation d'évolution du réseau, parfois appelée équation d'état: $M' = M + W \bar{\sigma}$. Le passage d'un marquage M à M' par le tir de σ est noté $M[\sigma \triangleright M']$. Un vecteur $\bar{\sigma}$ peut représenter plusieurs séquences de tirs différentes. Avant d'utiliser l'équation d'évolution il faut s'assurer qu'il existe au moins une séquence de tirs qui est effectivement tirable.

Nous désignons par $R(N, M_0)$ (respectivement $R(N, M)$) l'ensemble des états atteignables et par $L(N, M_0)$ (respectivement $L(N, M)$) l'ensemble des séquences tirables à partir de M_0 (respectivement M). Si l'ensemble $R(N, M_0)$ est fini (respectivement infini), il est possible de lui associer le graphe des marquages noté $G_m(N, M_0)$ (respectivement le graphe de couverture noté $G_c(N, M_0)$).

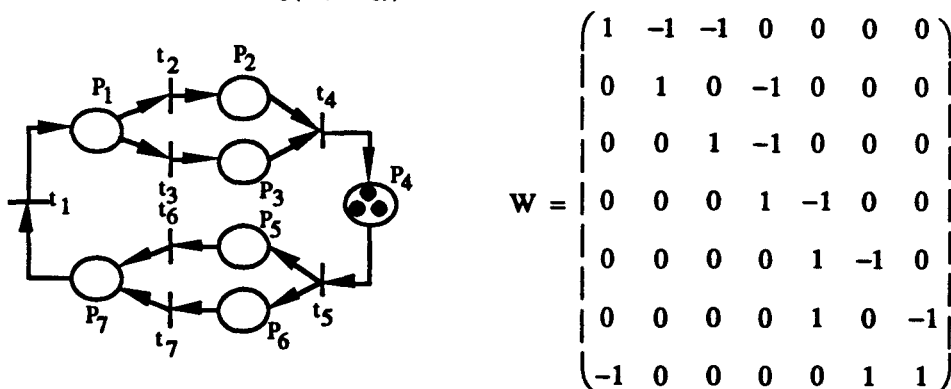


Fig. II-12 Exemple d'un RdP et de sa matrice d'incidence associée

II.4.2 Propriétés des RdP et caractérisation structurelle

En ce qui concerne les propriétés des RdP, une distinction peut être faite entre les propriétés structurelles et les propriétés comportementales. Les premières ne dépendent que de la structure du réseau, tandis que les secondes sont liées à un certain marquage.

Une transition est *vivante* si pour tout marquage accessible, il existe une séquence de tirs qui mène à un marquage à partir duquel cette transition est tirable, i.e. $\forall M \in R(N, M_0), \exists M' \in R(N, M)$, tel que $\forall p \in {}^\circ t, M'(p) \geq \text{Pré}(p, t)$. Un RdP est vivant si toutes ses transitions sont vivantes. Un RdP est structurellement vivant s'il existe un marquage M_0 qui rend le réseau vivant. Une transition est *bloquée* si elle n'est pas tirable. Un RdP est bloqué si aucune transition n'est tirable.

Une place p est *bornée*, s'il existe un entier $k \in \mathbb{N}$ tel que $M(p) \leq k, \forall M \in R(N, M_0)$. Un réseau est borné, si toutes ses places sont bornées. Un réseau est *sauf* si $k = 1$ pour toutes les places. Un réseau est structurellement borné s'il est borné pour tout marquage initial M_0 . Un marquage M' est un *état d'accueil*, s'il est accessible à partir de tout marquage M lui-même accessible à partir de M_0 , i.e. $M' \in R(N, M) \forall M \in R(N, M_0)$. Si M_0 est un état d'accueil, alors le réseau est dit *réinitialisable*.

L'analyse des RdP en vue de corroborer des propriétés citées ci-dessus peut s'effectuer sur le graphe des marquages (ou le graphe de couverture si le réseau n'est pas borné), par transformation (réduction, décomposition), par simulation ou par analyse structurelle. Ces approches sont généralement complémentaires et une analyse peut par exemple consister à transformer le RdP en un RdP équivalent mais de plus petite taille et pour lequel les états peuvent être énumérés. Ces quatre approches s'appliquent également pour l'analyse des performances si le RdP est temporisé. Dans la suite nous allons présenter quelques éléments de l'analyse structurelle.

Un vecteur $Y \in \mathbb{N}^n$ (respectivement $X \in \mathbb{N}^m$) est un P-semiflot ou P-invariant (T-semiflot ou T-invariant), si et seulement si $Y^T W = 0$ et $Y \neq 0$ ($WX = 0$ et $X \neq 0$). Un P-semiflot Y (T-semiflot X) est dit *canonique* si le pgcd de ses composantes est égal à un. Un P-semiflot Y (T-semiflot X) est dit *minimal* s'il n'existe pas d'autre P-semiflot Y' (T-semiflot X') qui vérifie* $Y' \leq Y$ ($X' \leq X$). L'ensemble des places d'un P-semiflot (transitions d'un T-semiflot) définit son support noté $\|Y\|$ ($\|X\|$). Un P-semiflot (T-semiflot) est à *support minimal*, si son support ne contient le support d'aucun autre P-semiflot (T-semiflot). Un P-semiflot (T-semiflot) est dit *élémentaire* s'il est canonique et à support minimal. Un RdP est dit *consistant* (respectivement *conservatif*) s'il existe un T-semiflot $X \geq (1, 1, \dots, 1)^T$ (respectivement P-semiflot $Y \geq (1, 1, \dots, 1)^T$). Nous désignerons par \mathcal{Y} l'ensemble des P-semiflots d'un RdP. Une *famille génératrice* est un ensemble de P-semiflots $\{Y_1, Y_2, \dots, Y_j, \dots, Y_k\}$ tel que pour tout vecteur $Y \in \mathcal{Y}$, il existe des coefficients $q_i \in \mathbb{Q}^+$ tels que $Y = \sum q_i Y_i$. Une famille génératrice est dite *minimale* et notée *f.g.m.*, si et seulement si $\forall j = 1..k, \{Y_1, Y_2, \dots, Y_j, \dots, Y_k\} / Y_j$ n'est pas génératrice. Etant donné que les coefficients q_i doivent être positifs, k peut être supérieur à la dimension de l'espace vectoriel engendré par \mathcal{Y} [TOU 85].

* nous supposons que l'inégalité s'applique à toutes les composantes.

Un certain nombre de propriétés se déduit alors de l'étude des invariants. Ainsi, un RdP conservatif est structurellement borné. De même, la consistance est la condition structurelle pour la vivacité d'un RdP. Dans la suite de ce mémoire nous allons introduire d'autres propriétés liées aux invariants pour une analyse quantitative.

II.4.3 Temporisation des RdP

L'intégration du facteur temps peut être effectuée par une temporisation des places ou des transitions. Si les places sont temporisées, la durée de temporisation correspond à un temps minimum de séjour dans une place. Dans cette approche les activités sont associées aux places et les transitions modélisent des événements de début ou de fin d'activité. Si les transitions sont temporisées, elles correspondent à des activités tandis que les places modélisent des états. L'équivalence entre ces différentes approches a été montrée et des règles simples de transformation ont été énoncées dans [SIF 80].

Dans la suite nous n'utilisons que des temporisations déterministes associées aux transitions du réseau définies par

$$\mu : T \mapsto \mathbb{IQ}^+ \quad \text{temporisations déterministes des transitions} \quad (\text{II.24})$$

Différentes sémantiques de tirs ont été définies et la plus courante est celle qui consiste à découper le tir d'une transition en *trois phases*. Si les conditions de franchissement de la transition t sont vérifiées, la première phase consiste à retirer le nombre de marques défini par $\text{Pré}(p,t)$ dans toutes les places $p \in \bullet t$. Ensuite, pendant la durée de tir $\mu(t)$, les marques disparaissent dans la transition. Dans une troisième phase, des marques, dont le nombre est spécifié par $\text{Post}(t,p)$, sont ajoutées dans les places $p \in t \bullet$. A cause de la disparition des marques pendant la phase d'activation de la transition, le bilan des marques n'est plus le même que pour le réseau non-temporisé. Pour la démonstration de certains résultats présentés dans la suite, une description plus fine de l'évolution du réseau a été nécessaire.

Une autre sémantique est celle du *tir atomique* où les marques sont gelées dans les places d'entrée pendant la durée de tir. A la fin de la durée de tir les marques sont enlevées des places d'entrée et ajoutées aux places de sortie comme spécifié par les

fonctions Pré et Post. L'avantage de cette deuxième sémantique est sa capacité de modéliser la préemption, c'est-à-dire l'interruption d'une activité entamée par une autre activité de plus haute priorité.

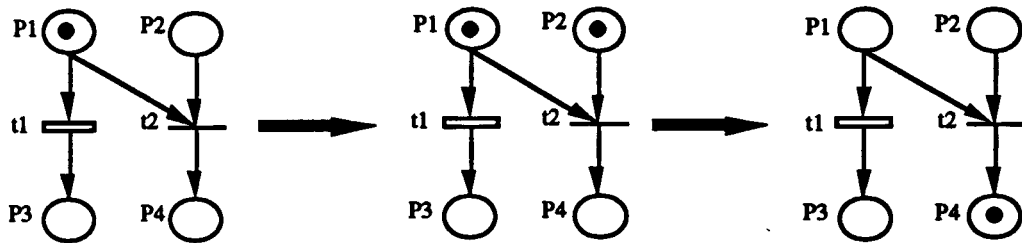


Fig. II-13 *Préemption du tir d'une transition temporisée*

Les deux sémantiques sont équivalentes si la modélisation des conflits est effectuée avec des transitions immédiates. Dans ce mémoire nous adoptons la sémantique de tir atomique, bien que nous n'utilisions pas la préemption dans nos modèles, mais parce qu'elle correspond mieux à la notion de *lieu* de transformation, utilisée dans le projet CASPAIM (cf. III.3.2). Remarquons que pour les deux sémantiques, la définition complète d'un état initial nécessite la prise en compte d'éventuels temps de tir résiduels.

Une source de confusion possible lors de l'analyse des RdP temporisés est liée au degré de ré-entrance des transitions. Certains auteurs interdisent d'office toute ré-entrance et supposent que les marques passent l'une après l'autre par la transition sans modéliser cette restriction explicitement par des boucles de capacité. Cette convention constitue cependant une limitation de la puissance de modélisation parce que les délais purs qui ne sont associés à aucune ressource ne peuvent plus être modélisés. Nous adoptons donc une sémantique de serveur infini et représentons, si cela est nécessaire, une limitation du degré de ré-entrance par une boucle de capacité associée.

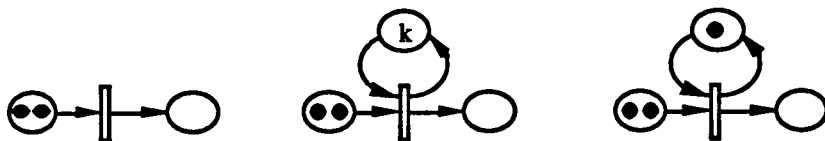


Fig. II-14 *sémantique de serveur infini, multiple ou simple*

II.4.4 Les graphes d'événements

Une classe particulière de Réseaux de Petri est constituée par les graphes d'événements. Dans les graphes d'événements chaque place possède exactement une transition d'entrée et une transition de sortie, c'est-à-dire $\text{Card}({}^{\circ}p) = \text{Card}(p^{\circ}) = 1$. De plus les poids de tous les arcs sont égaux à 1. Les graphes d'événements sont donc ordinaires et sans conflit au niveau de leur structure.

Un graphe d'événements est *fortement connexe*, s'il existe un chemin orienté de tout sommet (place ou transition) vers tout autre. Dans la suite nous ne considérons que des graphes d'événements fortement connexes. Un *circuit* dans un graphe d'événements est un chemin qui relie un sommet (place ou transition) à lui-même. Un *circuit élémentaire* est un circuit qui ne passe pas plus d'une fois par le même sommet.

Pour un graphe d'événements GE nous désignons par

$\gamma = (t_i, p_i, t_j, p_j, \dots, t_k, p_k)$	un circuit élémentaire de GE
$\Gamma = \{\gamma / \gamma \in GE\}$	l'ensemble des circuits élémentaires de GE
$M(\gamma) = \sum_{p_i \in \gamma} M(p_i)$	la somme des marques dans les places de γ
$\mu(\gamma) = \sum_{t_i \in \gamma} \mu(t_i)$	la somme des temporisations des transitions de γ
$C(\gamma) = \frac{\mu(\gamma)}{M(\gamma)}$	le temps de cycle associé à γ
$S_t(k)$	la date de début du $k^{\text{ème}}$ tir de la transition t

Les théorèmes suivants ont été énoncés dans [COM 71]. Ils sont valables pour des graphes d'événements avec ou sans temporisation:

◆ Le nombre de marques dans chaque circuit élémentaire est constant,
i.e. $\forall M \in R(GE, M_0), \forall \gamma \in \Gamma, M(\gamma) = M_0(\gamma)$ (II.25)

◆ Un graphe d'événements est vivant si et seulement si chaque circuit élémentaire est marqué, i.e. GE vivant $\Leftrightarrow \forall \gamma \in \Gamma, M(\gamma) > 0$ (II.26)

Les graphes d'événements temporisés ont été étudiés dans [RAM 80] et les résultats suivants ont été établis:

- ◆ Le temps de cycle C_Γ d'un graphe d'événements est supérieur au temps de cycle du circuit le plus lent, i.e. $C_\Gamma \geq C^* = \max_{\gamma \in \Gamma} \{C(\gamma)\}$ (II.27)

- ◆ Si les dates des premiers tirs des transitions sont convenablement choisies, il existe un régime de fonctionnement *périodique* avec des dates de tir des transitions données par $\forall t \in T, \forall k \in \mathbb{N}, S_t(k+1) = S_t(1) + kC^*$ (II.28)

Le mode de fonctionnement de *tir au plus tôt* a été étudié dans [CHR 85]. Comme le nom l'indique, chaque transition est tirée dès que possible:

- ◆ Dans un mode de fonctionnement de tir au plus tôt, un régime permanent K -périodique est atteint après un temps transitoire fini, i.e. $\exists K \in \mathbb{N}, \exists k_0 \in \mathbb{N}, \forall t \in T, \forall k \geq k_0, S_t(k+K) = S_t(k) + KC^*$ (II.29)

Le rapprochement entre les circuits et les P-semiflots d'un graphe d'événements est évident. Les circuits élémentaires correspondent aux P-semiflots élémentaires du réseau et peuvent alors être déterminés par des algorithmes développés pour le calcul des invariants [MAR 82]. En ce qui concerne les T-semiflots, la propriété suivante nous sera utile dans la suite:

- ◆ Un graphe d'événements fortement connexe possède un T-semiflot élémentaire unique donné par $X = (1, 1, \dots, 1)^T$ (II.30)

II.4.5 Modélisation de systèmes flexibles avec des graphes d'événements

Puisque les graphes d'événements ne permettent pas de représenter des conflits, leur domaine d'application est naturellement limité. Les seuls indéterminismes existant dans un graphe d'événements sont relatifs aux dates de tir des transitions. Avec des temporisations déterministes et un mode de tir des transitions défini (*tir au plus tôt* ou

périodique) le fonctionnement d'un graphe d'événements est entièrement spécifié. La modélisation d'ateliers flexibles par des graphes d'événements suppose donc que tous les conflits représentatifs des flexibilités de ces systèmes aient été résolus. Au vue de la démarche allant d'une planification, via une programmation et une planification fine vers un ordonnancement, un modèle de graphe d'événements représente alors le *résultat* d'un processus d'élaboration d'une commande.

La modélisation par graphes d'événements fortement connexes est particulièrement adaptée pour représenter des fonctionnements *cycliques*, appelés aussi fonctionnements *répétitifs*. Dans ces modes de fonctionnement, la commande de l'atelier est spécifiée sur un horizon limité et ensuite répétée un certain nombre de fois. Evidemment, dès que la commande est modifiée, le modèle associé n'est plus valable.

Un modèle graphe d'événements représentatif d'un fonctionnement répétitif d'un atelier flexible a été présenté dans [DUB 83], s'appuyant sur des gammes bouclées et des circuits représentatifs du comportement des ressources. Dans cet article est également évoqué la possibilité de tenir compte de temps de set-up et de temps de transport en ajoutant des transitions temporisées pour les premiers dans les circuits modélisant les ressources et pour les deuxièmes dans les circuits correspondant aux gammes bouclées.

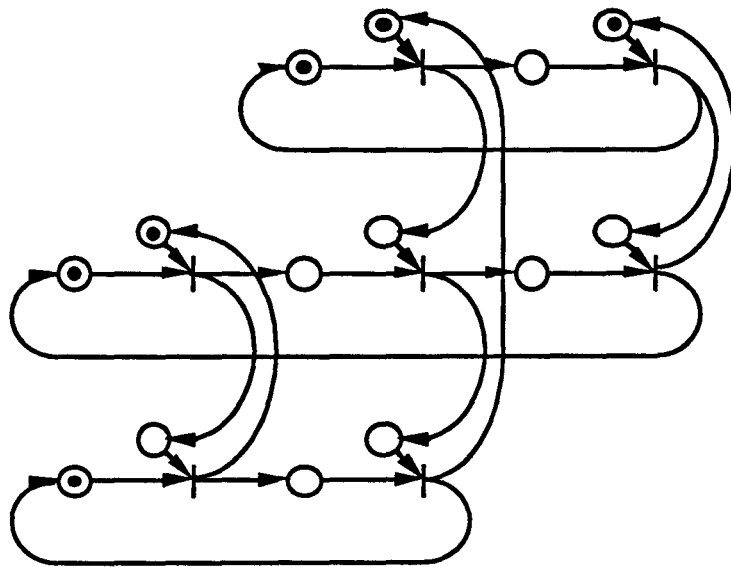


Fig. II-15 Graphe d'événements et fonctionnement répétitif [DUB 83]

Cette approche de modélisation a été poursuivie dans [HIL 89a][HIL 89b] en formalisant la démarche de construction du modèle graphe d'événements. L'approche part d'un ensemble de gammes de fabrication, représentant un ordre total entre les opérations. Ces gammes sont *dupliquées* autant de fois que nécessaire pour réaliser des ratios de production spécifiés (0.25, 0.25 et 0.5 pour l'exemple de la figure II-16).

Par le bouclage individuel des gammes (places r_i) des *circuits de fabrication* sont créés. Les marques dans ces circuits correspondent aux produits bridés sur leurs palettes. Les transitions qui représentent les opérations sont ensuite liées par des *circuits de commande* modélisant chacun une machine et l'ordre dans lequel cette machine traite les différents produits.

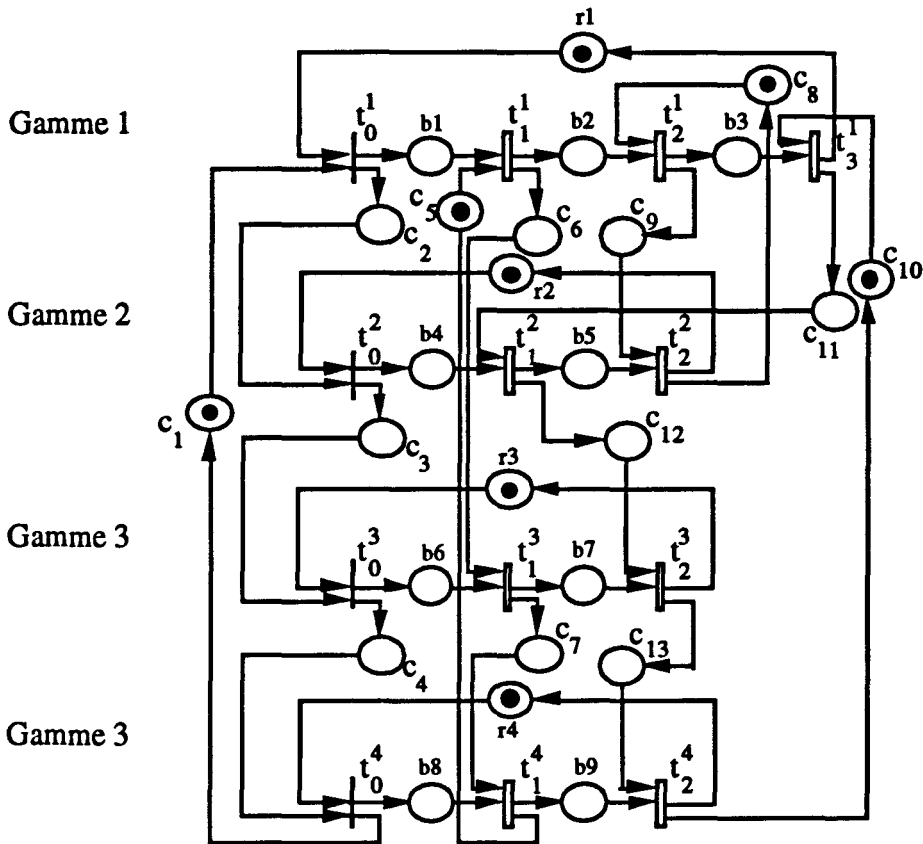


Fig. II-16 Modèle graphe d'événements d'un job-shop [HIL 89a]

En plus des circuits de fabrication et de commande, il existe un troisième type de circuits, appelé *circuits mixtes*, mis en évidence dans l'approche (par exemple $\gamma = (c_8, t_2^1, b_3, t_3^1, c_{11}, t_1^2, b_5, t_2^2)$) dans le réseau de la figure II-16). Ces circuits mixtes contiennent des places des deux autres types de circuits et résultent alors des choix des circuits de commande.

Les résultats concernant la vivacité (théorème II.26) et le temps de cycle (théorème II.27) des graphes d'événements s'appliquent donc également à ce type de circuits. L'absence de marques dans ces circuits mixtes permet alors d'expliquer les blocages de certaines configurations de la commande. De même, les conditions sur le nombre de marques nécessaires pour obtenir un temps de cycle imposé par le plus lent des circuits de commande portent alors sur les circuits de fabrication et les circuits mixtes.

Divers algorithmes de minimisation du nombre de marques dans un contexte de temporisations déterministes et stochastiques ont été proposés s'appuyant sur des modèles obtenus par l'approche de modélisation présentée [HIL 89b][LAF 91][LAF 92].

II.5 Conclusion

Dans la première partie de ce chapitre nous avons étudié la façon dont une conduite de flux spécifiée par des ratios de production peut s'intégrer dans une hiérarchie de gestion de production. Nous avons comparé deux approches de planification fine et en avons dégagé les caractéristiques fondamentales, notamment la façon dont les incertitudes dues aux pannes des machines sont traitées.

Les hypothèses sur les pannes et plus généralement sur les incertitudes sont d'une importance primordiale pour toute analyse quantitative. En effet, si les pannes sont trop fréquentes, l'élaboration d'une commande prévisionnelle ne peut plus être envisagée. Les répercussions de ces hypothèses se retrouvent au niveau de la modélisation et un modèle de type graphe d'événements n'est pas utilisable.

Dans la suite de ce mémoire, nous nous plaçons dans l'hypothèse de pannes de machines rares, justifiant l'existence d'un régime stable entre les événements caractérisant

des débuts ou fins de pannes. Cette façon d'appréhender le problème d'analyse du système implique alors une interprétation très restrictive du terme *commande* tel que nous l'utilisons pour désigner certains types de décisions.

En effet, nous supposons que la réactivité par rapport aux pannes de machines se situe à un niveau de hiérarchie *supérieur* à celui que nous considérons. De plus, dans le chapitre suivant, lors de la construction du modèle RdP et de l'analyse de la problématique relative au système de transport, nous allons mettre en évidence la nécessité de disposer d'une couche de commande *inférieure* intégrant une réactivité vis-à-vis de l'évolution asynchrone du processus physique réel.

Sous l'hypothèse d'un régime de fonctionnement stable, une phase d'évaluation préliminaire par flux peut alors être considérée comme le premier pas d'une commande déterministe. Dans la suite nous allons proposer une démarche d'optimisation progressive allant de l'évaluation préliminaire jusqu'à la construction d'un graphe d'événements représentatif d'un fonctionnement répétitif optimisé. Toute la démarche s'appuie sur une modélisation par RdP, outil dont nous avons justifié la pertinence pour la problématique étudiée.

III. Fonctionnement répétitif d'un atelier flexible: Modélisation par Réseaux de Petri et optimisation des flux de production

III.1 Introduction

Comme nous l'avons déjà indiqué, un graphe d'événements, en tant que modèle déterministe, peut seulement représenter le résultat d'un processus d'élaboration de la commande d'un atelier flexible. Il faut donc résoudre tous les indéterminismes représentatifs des flexibilités du système et notamment les indéterminismes au niveau des gammes possédant des ordres partiels ou les indéterminismes d'affectation si une opération peut être effectuée sur des machines différentes. Bien que les ordonnancements cycliques aient été étudiés dans la littérature, à notre connaissance, le problème de l'élaboration d'un ordonnancement cyclique n'a pas encore été abordé dans l'hypothèse d'une prise en compte simultanée de ces deux aspects particuliers.

Des ordonnancements cycliques ont été étudiés dans [HAN 89] dans le cadre d'une modélisation de boucles vectorielles. En raison de boucles récurrentes, des contraintes de précédence sont introduites entre les exécutions de tâches dans des itérations différentes. Ces contraintes sont spécifiques au contexte informatique et n'existent pas dans le domaine d'application des ateliers flexibles que nous considérons. Par ailleurs, le cas de ressources multiples a été pris en compte pour un ordonnancement cyclique dans [MUN 91]. Le problème d'affectation de ces ressources aux tâches est résolu en attribuant à chaque ressource toutes les tâches qu'elle est susceptible d'effectuer pendant une même itération.

Dans les travaux évoqués précédemment - faisant référence au contexte informatique - les notions de *ratio de production* et de *flexibilité de gamme* n'existent pas. Dans ce contexte, une modélisation par *graphes potentiel-tâches généralisé* [CHR 85] permet de modéliser des contraintes entre des itérations différentes sans qu'il soit nécessaire de disposer de l'équivalent des jetons d'un RdP. Par conséquent, *l'en-cours* n'est pas pris en compte explicitement, bien que la relaxation des contraintes de précédence entre les tâches, effectuée dans [MUN 91] pour réaliser un ordonnancement donné, y fasse implicitement appel.

Si l'on considère le domaine de la production flexible, l'approche de [HIL 89a] que nous avons présentée auparavant est basée sur l'hypothèse d'un ordre total entre les opérations et ne tient pas compte de groupes de machines identiques. Ces mêmes hypothèses sont utilisées dans [DUB 90] où une utilisation cyclique des palettes est étudiée. Dans [ERS 82] une approche de lancement cyclique de produits est présentée, supposant une discipline de PAPS (*premier arrivé premier servi*) au niveau des machines ainsi qu'un ordre total entre les opérations d'une gamme. Il est alors démontré qu'il est toujours possible de saturer la machine critique et une heuristique est présentée pour déterminer la séquence d'entrée qui minimise la taille du buffer au niveau de cette machine critique et, implicitement, la durée du régime transitoire.

Dans la suite, nous nous plaçons dans un contexte plus général en considérant des systèmes avec des gammes flexibles, des groupes de machines identiques, des ratios de production différents pour chaque type de produit et un ordonnancement a priori quelconque qui reste donc à déterminer pour chaque machine. Dans un premier temps nous précisons les indices de performances que nous associons au système et les facteurs ayant une influence sur ces indices. Nous en déduisons différents critères d'optimisation pour diverses formulations du problème d'optimisation. Afin de pouvoir aborder la complexité de la problématique retenue, nous nous appuyons sur une modélisation par RdP qui accompagnera tout le processus de résolution. Nous proposons une méthode systématique pour construire le modèle RdP et précisons ainsi la classe de systèmes qui est concernée par notre étude.

Dans le cas où des ratios de production sont donnés, nous effectuons une optimisation sur ce modèle en équilibrant au mieux la charge des machines. Nous en déduisons les performances maximales potentielles. La phase suivante d'analyse porte alors sur les conditions de réalisation de ces performances maximales. Nous établirons des conditions nécessaires et suffisantes pour leur atteignabilité et montrons que - bien que les performances maximales soient théoriquement atteignables - leur réalisation pratique est souvent impossible.

III.2 Fonctionnements répétitifs

III.2.1 Indices de performances

Un fonctionnement de type répétitif constitue une approche courante, bien que particulière dans le cadre d'une conduite d'atelier par commande prévisionnelle. Si les besoins de fabrication sont spécifiés sur un horizon long, l'idée de base est de limiter la complexité du problème d'élaboration de la commande en ne regardant dans un premier temps qu'une période d'une certaine durée. Il s'agit donc d'une première forme de réduction de la complexité évoquée par une *décomposition temporelle*. Tous les indéterminismes sont résolus sur cette période élémentaire, la commande est ensuite appliquée de manière répétitive, autant de fois que nécessaire pour produire les quantités de produits requises. Puisque la commande est identique pour chaque période élémentaire, les ratios de production doivent être respectés sur cette période. Nous supposons que les ratios sont donnés par

$$r_i = \frac{f_i}{\sum_i f_i} \text{ avec } f_i \in \mathbb{N} \text{ et } \text{pgcd}(f_i) = 1 \quad i = 1..I \quad (\text{III.1})$$

Il est alors naturel de définir l'ensemble minimal de produits qui satisfait des contraintes de ratio compatibles à la fois avec la production à effectuer et le caractère discret de la production ($f_i \in \mathbb{N}$).

Définition 1: L'ensemble minimal de produits, noté E_1 , est le plus petit ensemble contenant les produits dans les proportions spécifiées par les ratios de production.

De la définition des ratios (III.1) il suit que $\text{Card}(E_1) = \sum_i f_i$

Définition 2: Un cycle élémentaire de fonctionnement, noté CEF, est un fonctionnement qui permet de produire exactement les produits contenus dans E_1 en ramenant le système à son état initial.

A un CEF est alors associée une commande du système. Dans le cadre d'une modélisation par RdP nous pouvons représenter cette commande par un graphe

d'événements. Pendant un CEF chaque transition du graphe d'événements est tirée exactement une fois. Nous adoptons la règle de tir au plus tôt à l'intérieur d'un cycle et nous supposons que tous les événements à l'intérieur d'un cycle sont définis par rapport à une date $t_0 = 0$ à laquelle aucune des transitions n'est engagée dans un tir*. En utilisant les notations de la section II.4.4, le temps de cycle d'un CEF est alors défini par la plus grande date de fin de tir de toutes les transitions

$$C^1 = \max_{t \in GE} \{ S_t(1) + \mu(t) \} \quad (\text{III.2})$$

Ensuite, pour le fonctionnement répétitif, les dates de début de tir des transitions sont données par

$$S_t(k+1) = S_t(1) + kC^1 \quad k \in \mathbb{N} \quad (\text{III.3})$$

Le graphe d'événements de la figure III-1 modélise le fonctionnement cyclique de deux machines, M1 et M2, produisant deux types de pièces dans des proportions identiques. Le modèle contient quatre circuits, donnés par $\gamma_1 = (P2, t2, P8, t5, P6, t6, P7, t1)$, $\gamma_2 = (P4, t4, P9, t3)$, $\gamma_3 = (P1, t1, P2, t2, P3, t3, P4, t4)$ et $\gamma_4 = (P5, t5, P6, t6)$, associés respectivement à M1, M2, la gamme 1 et la gamme 2.

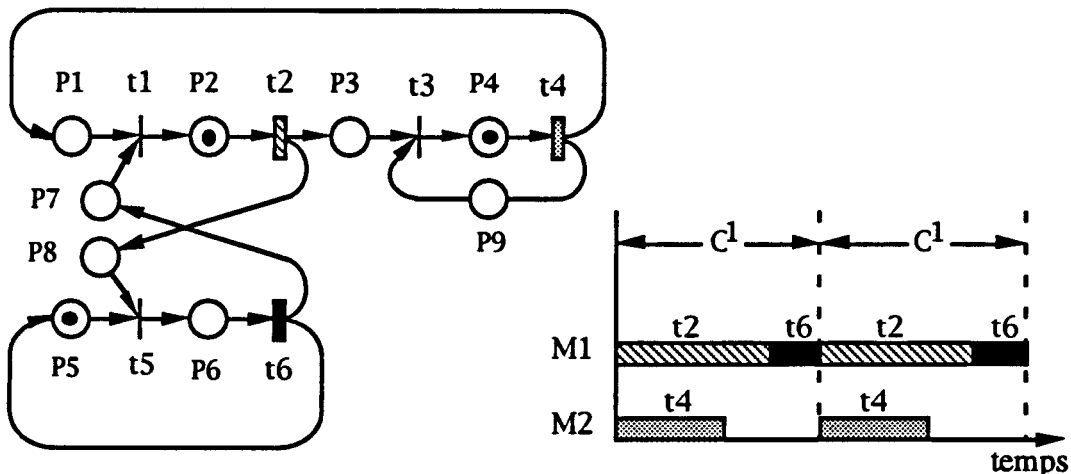


Fig. III-1 Graphe d'événements et fonctionnement répétitif

* l'absence de chevauchement entre cycles constitue évidemment une restriction pour les solutions envisageables pour l'ordonnancement mais ne restreint en rien la démarche proposée jusqu'à l'obtention du modèle ordonnançable (IV.2.3).

S'il y a plusieurs marques dans un circuit, pendant un CEF chaque marque avance jusqu'à la position initiale de la marque qui la précède directement. Dans l'exemple de la figure III-1 pendant un cycle de fonctionnement, la marque en P2 passe en P4 et celle de P4 en P2. Les termes *fonctionnement répétitif* et *retour à l'état initial* se justifient donc seulement parce que les marques ne sont pas individuellement identifiables dans notre modèle RdP. La présence de plusieurs marques dans un circuit peut alors être interprétée comme une relaxation des contraintes de précédence entre les tirs des transitions de ce circuit. Dans l'exemple, les tirs de t2 et t4 peuvent se chevaucher, ce qui serait impossible avec une seule marque dans le circuit γ_3 .

Définition 3: Un *cycle complexe de fonctionnement*, noté CCF, est un fonctionnement pendant lequel un multiple de E_1 est fabriqué en ramenant le système à son état initial et qui ne peut pas être décomposé en plusieurs CEF identiques.

Définition 4: L'*horizon*, noté v , associé à un CCF est le rapport entre le nombre de produits fabriqués pendant ce CCF et la cardinalité de E_1 :

$$v = \frac{\text{nombre de produits fabriqués pendant le CCF}}{\sum_i f_i}$$

L'ensemble de produits fabriqués pendant un CCF est noté E_v . Pour un CCF nous adoptons les mêmes hypothèses que pour un CEF. Nous supposons donc qu'il existe une date de début de cycle à laquelle aucune transition n'est engagée et qu'une sémantique de tir au plus tôt est appliquée à l'intérieur d'un cycle. La durée d'un CCF est alors notée C^v . Afin de rendre un CCF et un CEF comparables, nous introduisons la notion de *durée de cycle élémentaire virtuel* avec

$$C_v = \frac{C^v}{v} \quad (\text{III.4})$$

Il est évident que $C^1 = C_1$ mais généralement $C_v \leq C_1$. Une inégalité stricte signifie alors que l'allongement de l'horizon permet de réduire le temps de cycle et donc d'augmenter le débit de production. Le débit total de produits et le débit pour les produits de type i pour un fonctionnement cyclique (CEF ou CCF) sont donnés par

$$\Phi = \frac{\sum_i f_i}{C_v} = \frac{v \sum_i f_i}{C^v} \quad (\text{III.5})$$

$$\Phi_i = r_i \Phi = \frac{f_i}{C_v} = \frac{v f_i}{C^v} \quad i = 1..I \quad (\text{III.6})$$

Outre l'horizon v et le débit Φ , un troisième indice de performance d'un fonctionnement cyclique est donné par le niveau de l'en-cours. Ce facteur joue un rôle clé dans l'analyse de tout système de production, qu'il soit flexible ou non, soumis à un fonctionnement répétitif ou non.

Définition 5: L'en-cours de produits de type i , noté \bar{n}_i , est le nombre de produits de ce type qui se trouvent dans l'atelier de fabrication. L'en-cours total, noté \bar{n} , est la somme des en-cours par type de produit, soit $\bar{n} = \sum_{i=1..I} \bar{n}_i$.

Il est clair que l'en-cours varie constamment dans le type de système que nous considérons. En effet, l'en-cours est modifié par chaque chargement et déchargement au poste d'entrée/sortie de l'atelier. L'en-cours est donc une fonction du temps et nous avons en toute rigueur $\bar{n}_i = \bar{n}_i(t)$ et $\bar{n} = \bar{n}(t)$. Pour pouvoir tracer avec précision le niveau de l'en-cours, il faudrait connaître les dates exactes d'entrée et de sortie des pièces dans le système - ce qui est impossible tant qu'un ordonnancement n'a pas été défini. De toute façon la valeur instantanée de cet indice a une importance limitée.

En revanche, la connaissance de la valeur moyenne de l'en-cours peut être intéressante pour deux raisons. Premièrement, connaissant le débit de l'atelier, cette valeur peut être reliée au temps moyen de séjour de chaque produit dans l'atelier selon une relation de type *loi de Little*. Deuxièmement, la valeur moyenne de l'en-cours peut être utilisée pour une évaluation du capital moyen immobilisé. L'intérêt de ce dernier point dépend fortement de la valeur des produits et ne sera pas pris en compte dans notre analyse.

Plus intéressante que la valeur instantanée et la valeur moyenne de l'en-cours il faut considérer sa valeur maximale. La connaissance de cette valeur maximale est indispensable pour toutes les questions relatives au dimensionnement des zones de

stockage et du système de transport (qui peut être considéré comme un lieu de stockage mobile, voir III.3.2). Dans la suite nous allons considérer les bornes supérieures des valeurs instantanées des en-cours, notées n_i pour chaque type de produit et n pour l'en-cours global, et employer le terme *en-cours* dans ce sens par abus de langage.

III.2.2 Classes de décisions

Pour une démarche allant d'un niveau de planification fine jusqu'à la détermination d'une commande déterministe, nous avons regroupé les décisions en rapport avec les indéterminismes en classes distinctes. Certaines de ces classes existent dans tout système de production, tandis que d'autres sont spécifiques aux ateliers flexibles.

Classe D1: Choix des produits à fabriquer simultanément

Comme nous l'avons constaté lors de la comparaison de différentes approches de planification fine, le choix des produits peut être imposé par des contraintes extérieures. Il est également possible que des contraintes de faisabilité, liées par exemple aux capacités limitées des magasins d'outils, interdisent certaines combinaisons de types de produits. Il est clair que le choix de produire simultanément tels ou tels autres types de produits a une influence capitale sur les performances du système.

Classe D2: Choix des ratios de production

Les valeurs des ratios de production peuvent être soumises à des contraintes extérieures, par exemple pour tenir compte d'une demande dépendante. D'une manière générale, les ratios doivent répondre aux besoins de fabrication, mais dans le cadre d'une approche de type *regroupement dynamique*, les ratios de production peuvent également être utilisés pour équilibrer la charge des machines. De ce fait, ils influencent directement les performances du système.

Classe D3: Résolution des flexibilités de gamme

Cette classe peut être décomposée en trois sous-classes, D3a, D3b et D3c, en rapport avec trois types de flexibilité différents.

Le premier type de flexibilité existe si les gammes contiennent un ordre partiel entre les opérations. Dans ce cas nous parlons de *flexibilité de permutation* et affectons les décisions en rapport à la sous-classe D3a. Le deuxième type de flexibilité existe si une même opération peut être effectuée sur des ressources de types différents. Nous parlons alors de *flexibilité d'affectation* et regroupons les décisions en rapport dans la classe D3b. Un troisième type de flexibilité représente la possibilité de substituer une opération ou une suite d'opérations par une autre opération ou une autre suite d'opérations, éventuellement sur des ressources de types différents. Nous l'appelons *flexibilité de procédé* et affectons les décisions en rapport à la classe D3c.

Nous pouvons déjà signaler que les flexibilités de permutation n'ont pas d'influence sur la charge moyenne des machines pendant un calcul de flux. En revanche, les charges des différents types de machines sont évidemment influencées par la façon de résoudre les indéterminismes en rapport avec les classes D3b et D3c.

Classe D4: Affectation d'une opération à une ressource particulière

Après l'affectation d'une opération à un type de ressource, chaque opération doit être affectée à une ressource particulière s'il existe plusieurs ressources du même type.

Classe D5: Ordonnancement

Cette classe représente la problématique de déterminer le séquençement des produits sur les ressources. Dans le chapitre IV nous précisons un autre type de décisions associé à la classe D5 dans le contexte d'un fonctionnement répétitif. Remarquons qu'en travaillant sur un modèle RdP, le résultat de cette étape est un graphe d'événements.

Classe D6: En-cours

Cette classe comporte deux sous-classes. Il s'agit d'un côté de déterminer le niveau de l'en-cours pour chaque type de produit (sous-classe D6a) et de l'autre côté de décider de sa répartition dans le système (sous-classe D6b). Dans le contexte d'une modélisation par RdP il s'agit de déterminer le marquage du réseau.

Cette présentation des classes de décisions appelle plusieurs commentaires:

Premièrement nous voyons que l'en-cours peut être considéré d'une part comme un indice de performance et d'autre part comme une donnée qui fait partie de la commande du système et qui doit être fixée pour obtenir une commande déterministe. Nous avons donc une influence directe sur cet indice de performance, tandis que les deux autres indices, l'horizon et le débit, ne sont pas 'accessibles' directement. Ils apparaissent comme des conséquences plus ou moins lointaines des choix effectués. Remarquons que l'aspect commande de l'en-cours comporte son niveau et sa répartition, tandis que l'aspect indice se limite à son niveau. Deuxièmement, l'ordre dans lequel ces classes ont été présentées peut paraître naturel, il n'est cependant pas imposé. Il est toujours possible d'envisager une résolution simultanée de plusieurs classes successives, voire de toutes les classes à la fois. Pour des systèmes de tailles réelles, une démarche progressive est cependant recommandée pour maîtriser la complexité du problème.

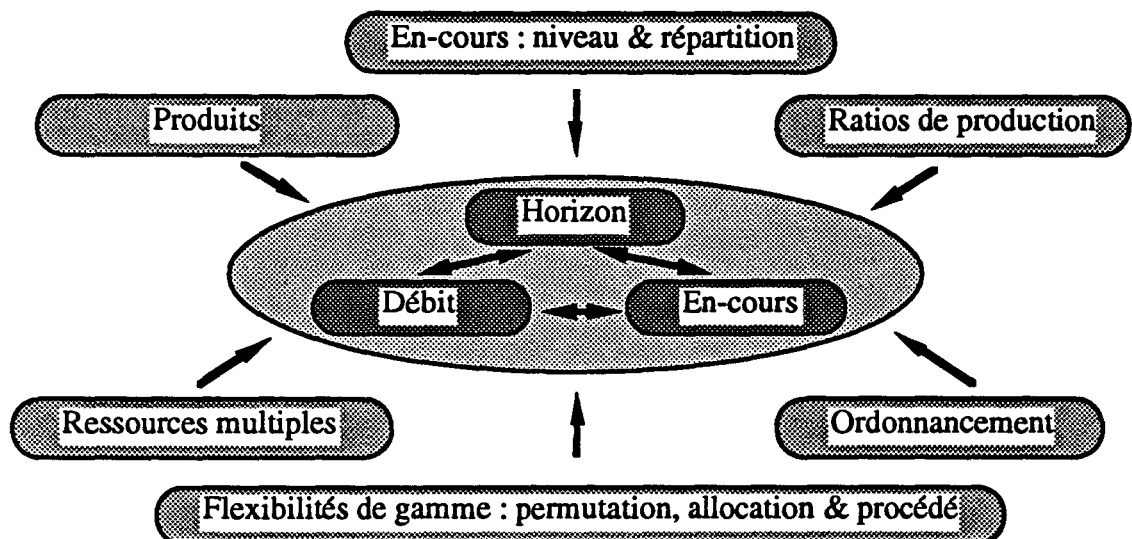


Fig. III-2 Indices de performances et facteurs d'influence

III.2.3 Formulation du problème

Après avoir spécifié les indices de performances et les classes de décisions prises en compte dans notre démarche, il convient maintenant de formuler avec plus de précision le problème à résoudre. Les trois indices peuvent être inclus dans le problème, selon le cas, sous forme de *critère d'optimisation* ou sous forme de *contrainte* à respecter.

	critère	contrainte
débit	maximiser Φ	$\Phi \geq \Phi_0$
horizon	minimiser v	$v \leq v_0$
en-cours	minimiser n	$n \leq n_0$

Fig. III-3 Formulation du problème

Il existe donc a priori $2^3 = 8$ formulations différentes du problème. Sept d'entre elles comportant au moins un critère d'optimisation sont des vrais problèmes d'optimisation. La formulation faisant intervenir uniquement des contraintes ($\Phi \geq \Phi_0$, $v \leq v_0$, $n \leq n_0$) est un problème d'analyse de faisabilité qui ne possède éventuellement pas de solution si les contraintes s'avèrent trop fortes. Remarquons que l'en-cours est considéré ici d'une manière globale. Il est néanmoins possible que l'en-cours de certains types de produits soit soumis à des contraintes tandis que celui d'autres types de produits soit à optimiser. Si l'en-cours de tous les I types de produits est pris en compte séparément, le nombre de formulations est égal à 2^{2+I} .

Certaines de ces formulations sont multi-critères et la façon dont ces critères sont effectivement pris en compte engendre différentes variantes du problème. En effet, plusieurs critères peuvent être optimisés conjointement, par exemple en introduisant une fonction de coût avec une pondération des critères. Une autre façon d'aborder un problème multi-critères consiste à introduire une hiérarchie stricte pour optimiser les critères successivement.

La formulation que nous avons retenue dans ce mémoire consiste à optimiser les trois indices de performances. Pour résoudre la complexité d'une optimisation conjointe de ces trois critères, nous traitons d'abord le conflit débit - horizon sous l'hypothèse d'un en-cours suffisant et nous optimiserons l'en-cours dans une deuxième phase. Cette formulation paraît intéressante parce que la connaissance du débit maximal est d'une grande importance pour l'intégration de l'atelier dans la gestion de production. En effet, toute la phase de programmation et de planification fine repose sur cet indice.

L'optimisation de l'horizon conjointement avec le débit se justifie par le fait qu'un horizon trop grand puisse interdire la réalisation effective du débit maximal. En effet, pour pouvoir employer une approche par fonctionnement répétitif, il faut être sûr de pouvoir répéter effectivement une commande. Mise à part l'influence fondamentale de l'horizon sur la faisabilité d'un fonctionnement répétitif, d'autres aspects doivent être pris en compte comme la complexité de la tâche d'ordonnancement et la durée du régime transitoire qui dépendent tous les deux de l'horizon.

Finalement, la séparation de l'optimisation du couple débit-horizon de l'optimisation de l'en-cours, justifiée dans ce contexte d'application, facilite une démarche progressive. En effet, la première optimisation peut être effectuée par un calcul de flux sur un modèle simplifié, évitant ainsi la complexité d'une approche combinatoire.

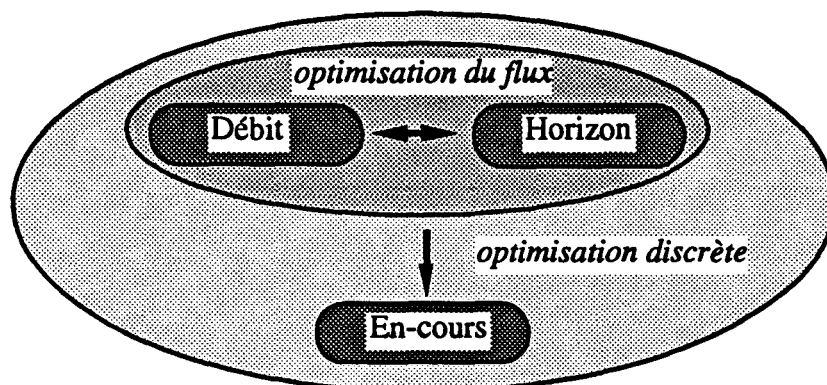


Fig. III-4 Hiérarchisation des critères d'optimisation

III.3 Construction du modèle Réseaux de Petri

III.3.1 Gammes de fabrication et ressources de transformation

Le développement du modèle RdP sur lequel repose notre optimisation est basé sur la connaissance des gammes de fabrication. Elles occupent en effet une place centrale dans les modèles utilisés dans le cadre du projet CASPAIM et nous nous appuyons ici sur les modèles et terminologies présentés dans [CRU 91] et [AMA 94].

Une *gamme logique* (GL) spécifie les opérations et l'ordre dans lequel celles-ci doivent être exécutées. La partie procédé n'est pas prise en compte et les GL ne spécifient donc ni les transferts, ni les ressources qui effectuent les opérations. Une flexibilité au niveau des GL modélise alors soit la possibilité d'une permutation de l'ordre des opérations (D3a), soit un choix entre des procédés de fabrication différents (D3c). Puisque nous ne considérons pas les assemblages dans ce mémoire, les GL ne comportent pas de synchronisations.

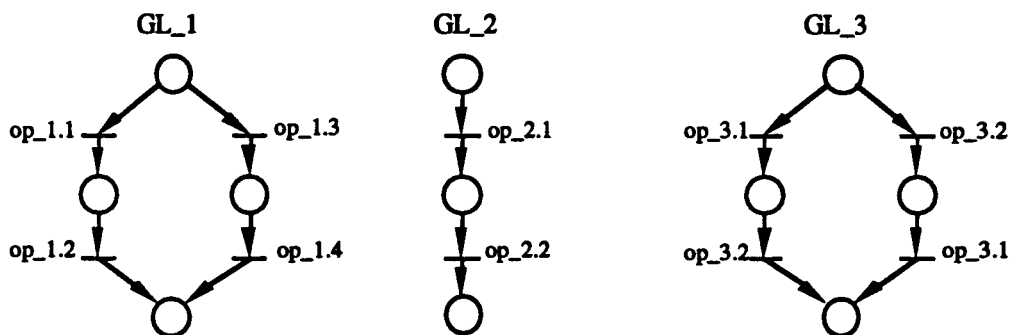


Fig. III-5 Gammes logiques

L'étape suivante de la démarche adoptée dans CASPAIM consiste à ajouter aux GL les informations relatives au procédé afin de spécifier sur quelle ressource une opération est effectuée. Le résultat est une gamme opératoire (GO) et la connaissance des ressources - et donc des lieux physiques - de transformation permet alors de spécifier les transferts entre ces ressources. L'affectation d'une opération à une ressource particulière nécessite cependant la résolution des décisions de la classe D4, décisions que nous ne pouvons pas résoudre à ce stade de l'analyse. En effet, tant que les flexibilités d'affectation (D3b) et de procédé (D3c) ne sont pas résolues, les opérations qui sont

définitivement affectées à un certain type de ressource ne sont pas connues et ne peuvent par conséquent pas être réparties entre les ressources de ce même type.

Nous proposons donc d'introduire des *gammes opératoires restreintes* (GOR) dans lesquelles toute opération est associée à un type de ressource plutôt qu'à une ressource identifiée. Si une opération peut être effectuée sur des ressources de types différents, alors cette flexibilité (D3b) apparaît dans les GOR comme le montre la transformation de GL_2 de la figure III-5 en GOR_2 de la figure III-6.

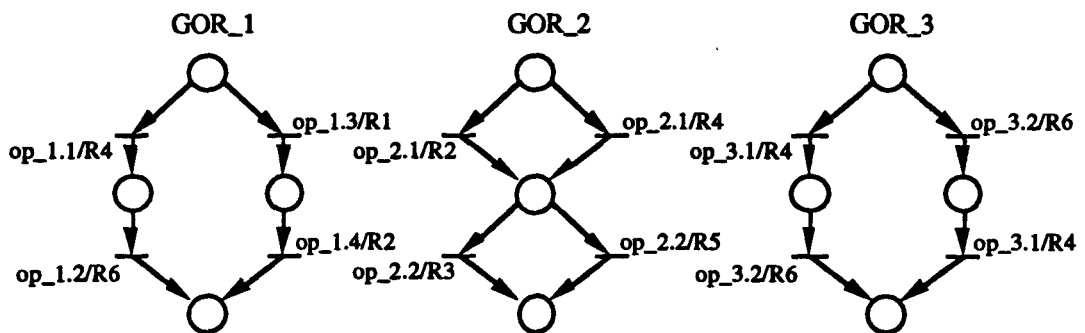


Fig. III-6 *Gammes opératoires restreintes*

Dans une GOR les trois types de flexibilités peuvent a priori exister plusieurs fois. Un premier type de problèmes apparaît lorsqu'une même gamme contient plusieurs flexibilités de procédé ou d'affectation enchaînées, comme c'est le cas de la GOR_2. Au lieu de modéliser deux conflits avec respectivement deux choix, nous pourrions modéliser cette situation sous forme d'un seul conflit avec quatre choix comme indiqué sur la figure III-7. Lors d'une optimisation de flux qui va tenter d'équilibrer les charges des machines, les flux qui traversent les transitions sont fixés et permettent alors de déterminer les taux de routages.

Si un calcul de flux, s'appuyant sur un modèle avec deux conflits contenant respectivement deux choix, indique par exemple que les quatre types de ressources doivent supporter des charges identiques, nous pouvons en déduire que le routage à respecter est $\alpha_1 = \alpha_2 = 0.5$. Il est alors possible de réaliser ce flux pour deux produits en routant un produit par R2-R3 et l'autre par R4-R5 ou, alternativement, un produit par R2-R5 et l'autre par R4-R3. Les deux possibilités sont strictement équivalentes en ce qui

concerne les charges moyennes induites sur les machines mais différent au niveau de l'association des résultats du premier choix avec ceux du deuxième choix.

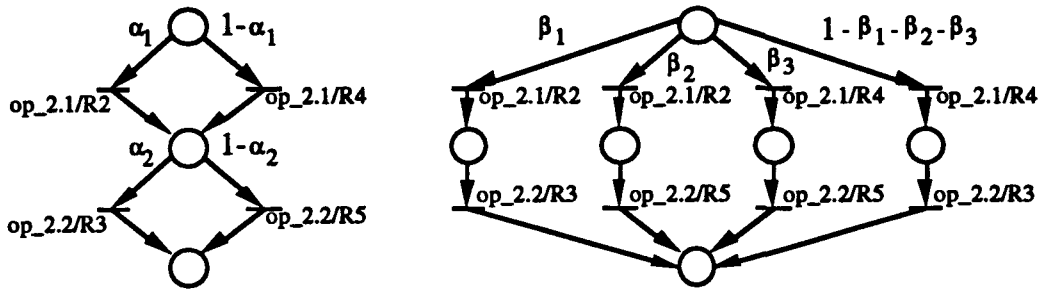


Fig. III-7 Modélisation de flexibilités enchaînées

En revanche, un calcul de flux s'appuyant sur une modélisation qui explicite les associations des choix aux différents niveaux, ne peut pas fixer de manière unique les paramètres β_1 , β_2 et β_3 . En effet, il existe plusieurs combinaisons pour réaliser un chargement identique des machines, par exemple ($\beta_1 = \beta_2 = \beta_3 = 0.25$) ou ($\beta_1 = \beta_3 = 0.5$ et $\beta_2 = 0$) pour n'en citer que deux. Pour la détermination des charges moyennes des machines les deux approches de modélisation sont équivalentes, mais puisque nous allons utiliser les valeurs numériques de ces taux de routage pour déterminer l'horizon v sur lequel ces taux peuvent être réalisés, nous modélisons des flexibilités enchaînées par des conflits successifs. Nous parlons alors de *modélisation pliée* des flexibilités enchaînées.

Un deuxième type de problèmes apparaît lorsque des flexibilités de type affectation ou procédé sont imbriquées comme dans l'exemple de la figure III-8.

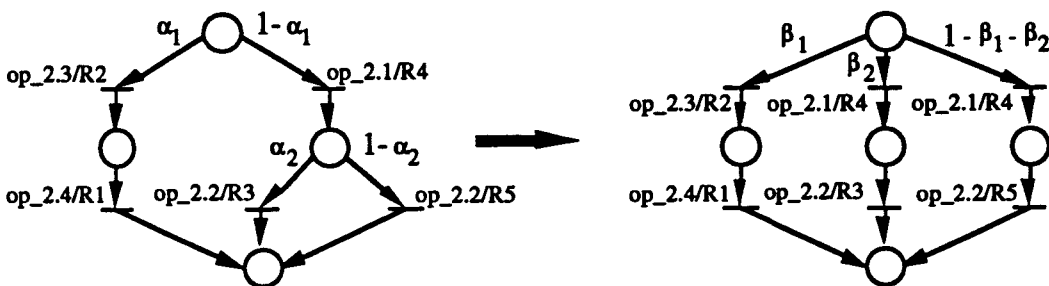


Fig. III-8 Flexibilités imbriquées

Dans la gamme de la figure III-8 qui représente une flexibilité de procédé contenant une flexibilité d'affectation dans une de ses branches, les taux de visite des transitions ne sont plus des fonctions linéaires des α_i , mais dépendent également du produit $\alpha_1\alpha_2$. Dans le cas particulier de la figure III-8, une transformation est possible, parce qu'une seule flexibilité d'affectation est imbriquée dans la flexibilité de procédé. En présence de plusieurs flexibilités d'affectation ou de procédé enchaînées dans une branche d'une autre flexibilité, un *dépliage* engendrerait la problématique évoquée précédemment. Nous devons donc choisir entre les trois options suivantes:

- Maintenir une modélisation pliée et accepter des taux de visite non-linéaires.
- Déplier les flexibilités et accepter des taux de routage non uniques.
- Exclure de notre analyse des cas où des chaînes de flexibilités sont imbriquées.

Dans ce mémoire nous avons adopté la troisième option et nous ne considérons donc pas les cas des chaînes de flexibilités imbriquées. Par conséquent, les taux de visite des transitions resteront linéaires en α_i et la problématique concernant la non-unicité au niveau des valeurs des α_i n'existe pas.

En ce qui concerne les flexibilités de permutation, modélisées par exemple dans la GOR_3 de la figure III-6, nous supposons qu'elles ne concernent que des opérations successives et qu'elles ne contiennent pas de flexibilités d'affectation ou de procédé imbriquées. Puisque les choix effectués pour résoudre les flexibilités de permutation n'influencent pas les charges moyennes des machines, nous allons 'masquer' ces flexibilités pendant le calcul de flux en attribuant des valeurs constantes aux taux de routage associés. Par conséquent, le problème de non-linéarité évoqué ci-dessus ne se pose pas au niveau des taux de routage associés aux flexibilités de permutation et nous pouvons donc admettre sans difficulté la présence de flexibilités de permutation imbriquées dans d'autres types de flexibilités.

Nous complétons les GOR en remplaçant chaque transition représentative d'une opération de transformation par l'élément *transition immédiate/ place / transition temporisée*. Nous appelons ces gammes alors *gammes opératoires restreintes temporisées* (GORT). Cette modélisation est introduite pour assurer que le RdP restera pur après connexion des ressources aux gammes. De plus, nous modélisons les choix représenta-

tifs des flexibilités par des choix libres et associons les types de machines aux opérations. La figure III-9 présente un exemple illustratif de 3 produits et 6 types de ressources qui sera développé dans la suite de ce mémoire.

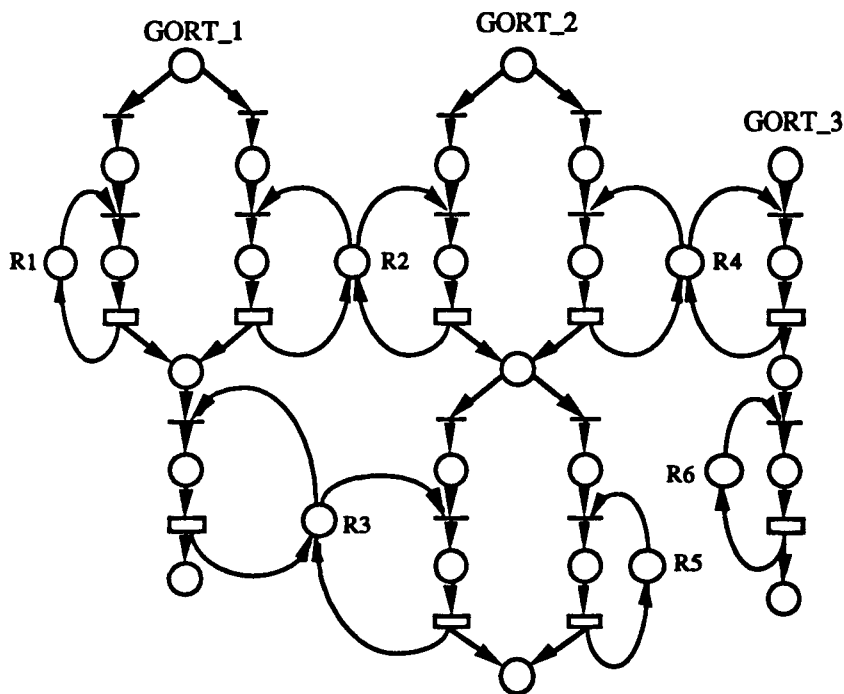


Fig. III-9 Types de ressources associés aux gammes opératoires restreintes temporisées (GORT)

III.3.2 En-cours et système de transport

Comme nous l'avons déjà évoqué, les transferts entre les lieux physiques où les opérations sont exécutées ne peuvent pas être détaillés tant que les opérations ont seulement été affectées à un type de ressources et non pas à une ressource particulière. De plus, comme conséquence de la modélisation pliée des flexibilités, les séquences d'opérations ne sont pas représentées de façon déterministe dans notre modèle si plusieurs flexibilités s'enchaînent.

A ce stade du développement du modèle il faut alors décider si les transferts entre les lieux physiques peuvent être négligés ou si des estimations des durées peuvent être incluses dans le modèle. Pour justifier le choix effectué dans ce mémoire nous détaillons

la modélisation d'exemples de deux types de systèmes de transport, d'un côté un convoyeur avec des palettes, de l'autre côté un système de chariots guidés automatiquement.

Dans le cas d'un système de transport réalisé par un convoyeur, nous supposons qu'il peut y avoir différents types de palettes dans le système. Ces palettes peuvent être dédiées à un type de produit particulier ou au contraire être universelles et capables de transporter des produits de types différents. Une première politique de gestion des palettes consiste à supposer qu'une palette vide est immédiatement rechargée avec une pièce brute et qu'elle reste affectée à cette pièce jusqu'à ce que la pièce quitte le système. Même si la pièce est dépalettisée pour un usinage, la palette lui reste réservée. Dans cette hypothèse de modélisation, l'en-cours et le nombre de palettes sont confondus et la limitation des deux peut être modélisée par le même bouclage.

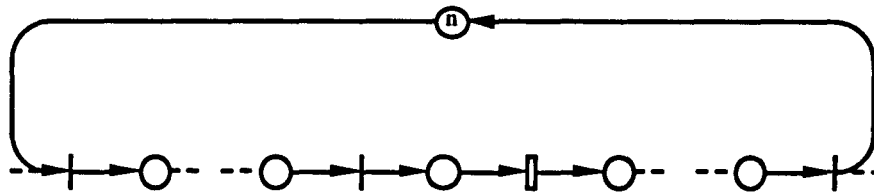


Fig. III-10 Bouclage commun pour l'en-cours et les palettes

Une autre politique de gestion des palettes consiste à libérer une palette pendant l'usinage de la pièce. Avec une telle politique de gestion, les palettes apparaissent explicitement comme des ressources de transport et l'en-cours doit être contrôlé séparément.

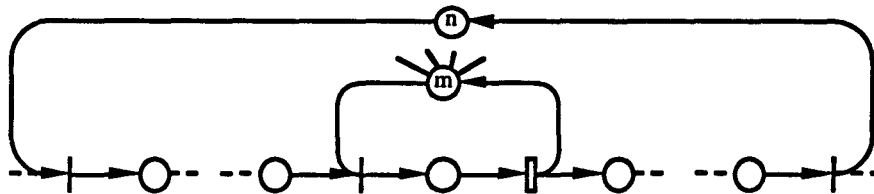


Fig. III-11 Contrôle séparé de l'en-cours et des ressources de transport

L'avantage de cette deuxième approche consiste en une meilleure utilisation possible des palettes et permet d'espérer un fonctionnement avec un nombre restreint de palettes. Cette utilisation plus intensive peut cependant induire des retards et ainsi nuire au

débit de production. En effet, puisque les palettes sont explicitement modélisées comme des ressources, elles peuvent devenir critiques et limiter le débit maximal de l'atelier.

Le modèle de la figure III-11 est également utilisable si les transports sont effectués par des chariots, généralement disponibles en nombre limité. Au niveau de la modélisation pour une analyse quantitative, la caractérisation de différents systèmes de transport dépend alors essentiellement de l'assimilation ou non de l'en-cours aux ressources de transport. Deux classes d'hypothèses sont donc à considérer:

- (a) les ressources de transport et l'en-cours sont confondus
- (b) les ressources de transport et l'en-cours sont modélisés séparément

Cette distinction n'a pas a priori de rapport avec le fait de négliger ou d'estimer les durées des transports. Il est néanmoins clair qu'une modélisation dans le cas (b), où les ressources de transport apparaissent explicitement dans le modèle, nécessite une prise en compte des temps de transport. Si les temps de transport sont négligés, les ressources de transport n'ont pas de charges associées et peuvent par conséquent être enlevées du modèle utilisé pour l'analyse quantitative.

Deux autres aspects qui doivent être pris en considération sont les zones de stockage intermédiaires et le système de manutention. Mis à part de rares cas où les pièces peuvent rester bridées sur les palettes pendant l'opération - par exemple pour des assemblages - la manutention des pièces est généralement effectuée par des robots. Dans le cadre du projet CASPAIM, la problématique relevant de la prise en compte de ces aspects a été étudiée dans [AMA 94]. La notion de lieu physique a été différenciée selon qu'il s'agisse d'un

- lieu mobile, désigné par T (ressource de transport)
- lieu actif, désigné par M (machine)
- lieu statique, désigné par S (stock)

Selon cette approche, il existe donc 9 types de transferts qui peuvent être modélisés de manière générique comme indiqué dans la figure III-12. Dans cette modélisation nous supposons d'une part que la ressource de manutention engage un transfert seule-

ment si le lieu d'accueil est libre et d'autre part que le lieu de départ est seulement libéré à la fin d'un transfert.

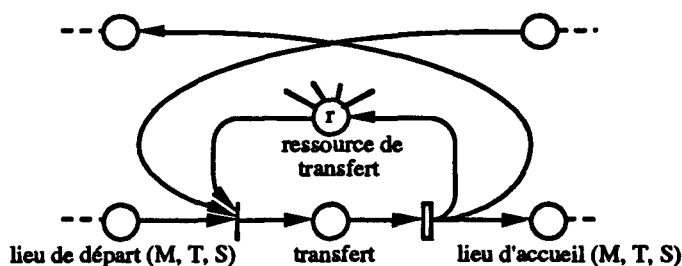


Fig. III-12 Transfert entre deux lieux physiques

Le transfert d'une pièce d'un lieu vers un autre peut alors

- (i) s'effectuer sans attente
- (ii) être retardé à cause de l'indisponibilité de la ressource de manutention
- (iii) être retardé parce que le lieu d'accueil n'est pas libre

En ce qui concerne (ii), une possibilité consiste à faire l'hypothèse que la ressource de manutention est toujours disponible. Cette hypothèse ne peut être validée qu'a posteriori, une fois que les opérations ont été associées aux machines et donc que les besoins de manutention sont connus. Les ressources de manutention peuvent également être incluses dans un calcul d'optimisation de flux et l'équilibrage des charges s'effectue alors en tenant compte de ces nouvelles ressources. Si le taux moyen d'utilisation d'une ressource de manutention est faible nous pouvons supposer que cette ressource est toujours disponible.

En revanche, si une ressource de manutention est critique ou si son taux moyen d'utilisation est élevé, il faut en tenir compte dans l'élaboration de l'ordonnancement. Dans ce cas se pose le problème d'un ordonnancement avec des opérations qui nécessitent la présence simultanée de plusieurs ressources imbriquées. Nous n'abordons pas cette classe de problèmes et considérons dans la suite l'hypothèse selon laquelle les ressources de manutention ne retardent pas les transferts. Nous supposons donc que la détermination de la commande des robots s'effectue à un niveau inférieur à l'ordonnancement, sans que celui-ci soit perturbé. Les durées de chargement et déchargement peuvent

néanmoins être prises en compte au niveau de l'ordonnancement en ajoutant une estimation de ces durées aux temps opératoires et aux temps de transport.

En ce qui concerne (iii), nous distinguons deux cas, selon que le lieu de départ est associé à une ressource modélisée explicitement et donc disponible en nombre limité ou non. Si le lieu de départ est un lieu statique ou un lieu mobile d'un système de transport de type (a), le transfert s'effectue seulement si un lieu d'accueil de type (M) ou (T) est disponible. En revanche, si le lieu de départ est un lieu actif ou un lieu mobile d'un système de transport de type (b), pour éviter des blocages, il faut que la pièce soit immédiatement évacuée à la fin de l'opération (transformation ou transport) associée au lieu. Par conséquent, nous devons supposer des zones de stockages de capacités suffisantes en aval des ressources disponibles en nombre limité.

Nous considérons donc les trois façons suivantes pour modéliser les transports:

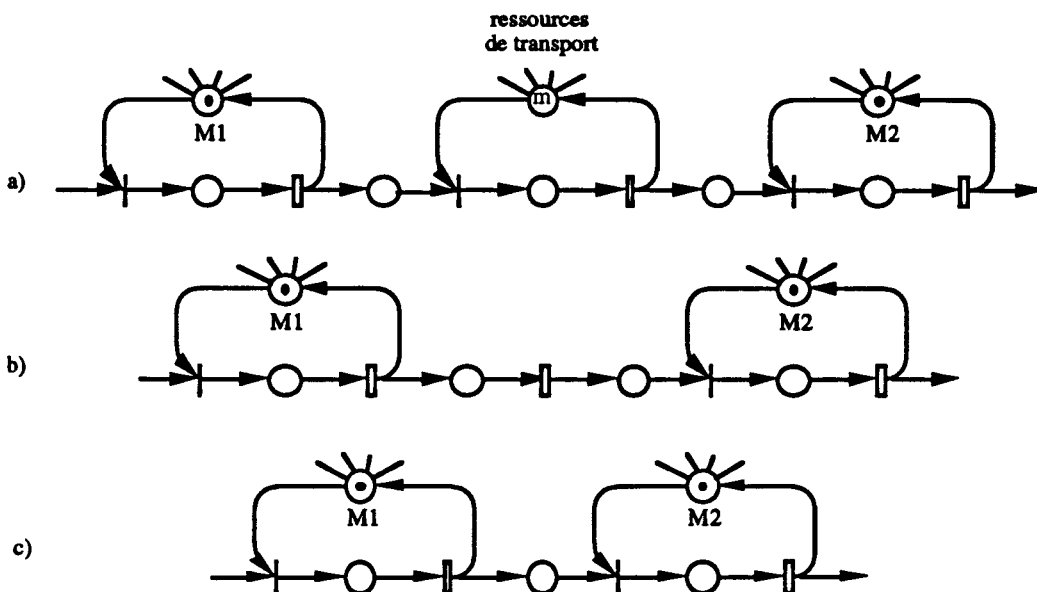


Fig. III-13 Modélisation des transports

Le cas a), modélisant un transport par une ressource de transport, et c), modélisant un transfert direct entre deux machines, rentrent directement dans le cadre de la modélisation proposée auparavant, seul le cas b) constitue une exception, car les transitions modélisant le transport n'ont pas de boucles de capacité associées. Par

conséquent, ces transitions ne jouent aucun rôle pendant un calcul d'optimisation qui consiste à équilibrer les charges des ressources. Les durées associées à ces transitions interviendront ultérieurement, lorsque l'en-cours sera considéré explicitement. Pour un calcul de flux nous pourrions donc a priori négliger ces durées.

Le degré de ré-entrance théorique maximal est cependant limité - par l'en-cours maximal par exemple - et afin d'homogénéiser la modélisation, nous pourrions associer une boucle de capacité 'suffisante', sans que cela ne change notre interprétation. Par conséquent, les trois types de transports se modélisent avec les primitives de bases introduites en III.3.1. Dans la suite nous ne distinguons plus les machines et les ressources de transport ainsi que les opérations de transformation ou de transport. Nous parlons de façon plus générale de ressources et d'opérations.

En ce qui concerne l'estimation des durées de transport, le problème est difficilement abordable sur le plan abstrait. En effet, ces durées dépendent fortement de l'agencement des machines et de la topologie du système de transport. Supposons qu'une durée de transport ait été estimée à une certaine valeur δ et que l'ordonnancement ait été construit en tenant compte de cette valeur. Si la durée effective est plus grande que δ , il est possible que l'opération suivante soit retardée et que ce retard soit 'irratrapable'. Dans ce cas nous parlons d'une durée de transport *apparente*. En augmentant l'en-cours, cette durée de transport peut être *masquée* et la durée apparente ramenée à zéro.

En revanche, si la valeur réelle a été surestimée et si la durée effective de transport est plus courte que prévue, la pièce arrivera en avance et devra attendre la disponibilité de la machine. Il apparaît alors assez naturellement qu'une surestimation systématique des durées de transport tend à accroître l'en-cours nécessaire, à l'inverse une sous-estimation risque de nuire au débit de production maximal. Ceci confirme notre idée de découpler l'optimisation du flux et de l'horizon de l'optimisation de l'en-cours.

En résumant l'analyse effectuée dans cette section, nous pouvons dire que la difficulté de la prise en compte des transports se situe à plusieurs niveaux:

(1) Si les robots de manutention sont fortement utilisés, ils sont susceptibles d'induire des retards et doivent alors être pris en compte explicitement dans l'élaboration d'un

ordonancement. A cause de l'imbrication de ces robots avec d'autres ressources le problème d'ordonancement se complique.

(2) Les transports entre les lieux où les opérations sont effectuées ne peuvent pas être explicités pour deux raisons. Premièrement, pour un calcul de flux, les opérations sont seulement affectées à des types de ressources dont les exemplaires individuels peuvent se trouver à des endroits différents. Deuxièmement, en présence de flexibilités enchaînées, les séquences d'opérations n'apparaissent pas explicitement dans une modélisation pliée.

(3) Les durées de transport dépendent fortement de l'architecture du système. Même si cette architecture est connue et le fonctionnement spécifié, il subsiste toujours des incertitudes sur les durées effectives (accrochage de palette, ...).

Ces remarques tendent à indiquer que la commande d'un système de production réel ne peut jamais être totalement déterminée au niveau de l'ordonancement. Il apparaît alors la nécessité d'un module de pilotage intégrant une forte réactivité de la commande vis-à-vis de l'évolution asynchrone du système physique [TAW 95]. D'une manière plus générale, la décomposition en trois niveaux *planification / ordonnancement / pilotage*, retenue dans le projet CASPAIM, se trouve justifiée du point de vue de l'analyse et donc de la maîtrise des performances du système.

III.3.3 Bouclage du modèle

Le bouclage de notre modèle s'effectue à deux niveaux et nous distinguons un *bouclage intérieur* et un *bouclage extérieur*. Le premier traduit la limitation de l'en-cours dans les gammes et peut, comme nous l'avons indiqué, également modéliser la limitation du nombre de palettes. Ce bouclage est indispensable, parce que nous supposons qu'il y a toujours des pièces brutes à l'entrée du système. En l'absence de bouclage, l'en-cours pourrait croître indéfiniment.

Nous supposons qu'il existe H types de contraintes sur l'en-cours (ou le nombre de palettes) et que chaque contrainte peut être associée à un ou plusieurs types de produits. Ceci permet la modélisation de contraintes additives qui seraient par exemple dues à des palettes pouvant supporter plusieurs types de produits différents. Pour

spécifier ces contraintes, nous associons à chaque contrainte $h = 1..H$ un ensemble S_h contenant les indices des types de produits concernés par cette contrainte. La contrainte h s'écrit alors

$$\sum_{i \in S_h} n_i \leq e_h \quad h = 1..H \quad (\text{III.7})$$

$$\text{avec} \quad \bigcup_h S_h = \{ 1..I \} \quad (\text{III.8})$$

$$\text{et} \quad \bigcap_h S_h = \emptyset \quad (\text{III.9})$$

La condition III.8 garantit que chaque type de produit est effectivement soumis à une contrainte tandis que la condition III.9 interdit l'existence de plusieurs contraintes sur le même type de produits. Dans l'exemple de la figure III-14 nous avons modélisé deux contraintes sur l'en-cours, la première concernant les types 1 et 2 (place P35), la deuxième concernant le type 3 (place P36).

Le deuxième type de bouclage concerne la totalité des gammes et - contrairement aux bouclages intérieurs - n'a pas d'interprétation physique directe. Il s'agit d'un artifice habituel en théorie des graphes permettant de lier les gammes au niveau de la structure du RdP. Dans l'exemple de la figure III-14, il existe des ressources partagées entre toutes les gammes, ce qui permettrait théoriquement d'imposer les ratios de production par l'affectation de ces ressources. De même, les ratios des gammes 1 et 2 pourraient être contrôlés par le conflit au niveau de la place P35, modélisant le bouclage commun de ces deux gammes. Puisque l'existence de ressources partagées ou de bouclages intérieurs communs ne peut pas être supposée dans le cas général, il est nécessaire d'introduire un bouclage commun qui permet de prendre en compte les ratios de production.

L'analyse des performances que nous proposons d'effectuer consiste dans un premier temps à déterminer les taux de visite des transitions du RdP. Ces taux sont définis comme étant le nombre de tirs des transitions par unité de temps et correspondent aux débits que nous cherchons à déterminer et - si possible - à optimiser. Afin de déterminer les taux de visite absolus, une première étape consiste à considérer les taux de visite relatifs, normalisés par rapport à une transition de référence, a priori quelconque. Nous avons choisi de normaliser les taux de visite par rapport à la transition de bouclage

commun, t_1 , dont un tir correspond au lancement en fabrication d'une pièce. Remarquons que cette façon de procéder est tout à fait analogue à celle employée pour analyser un réseau de files d'attente fermé.

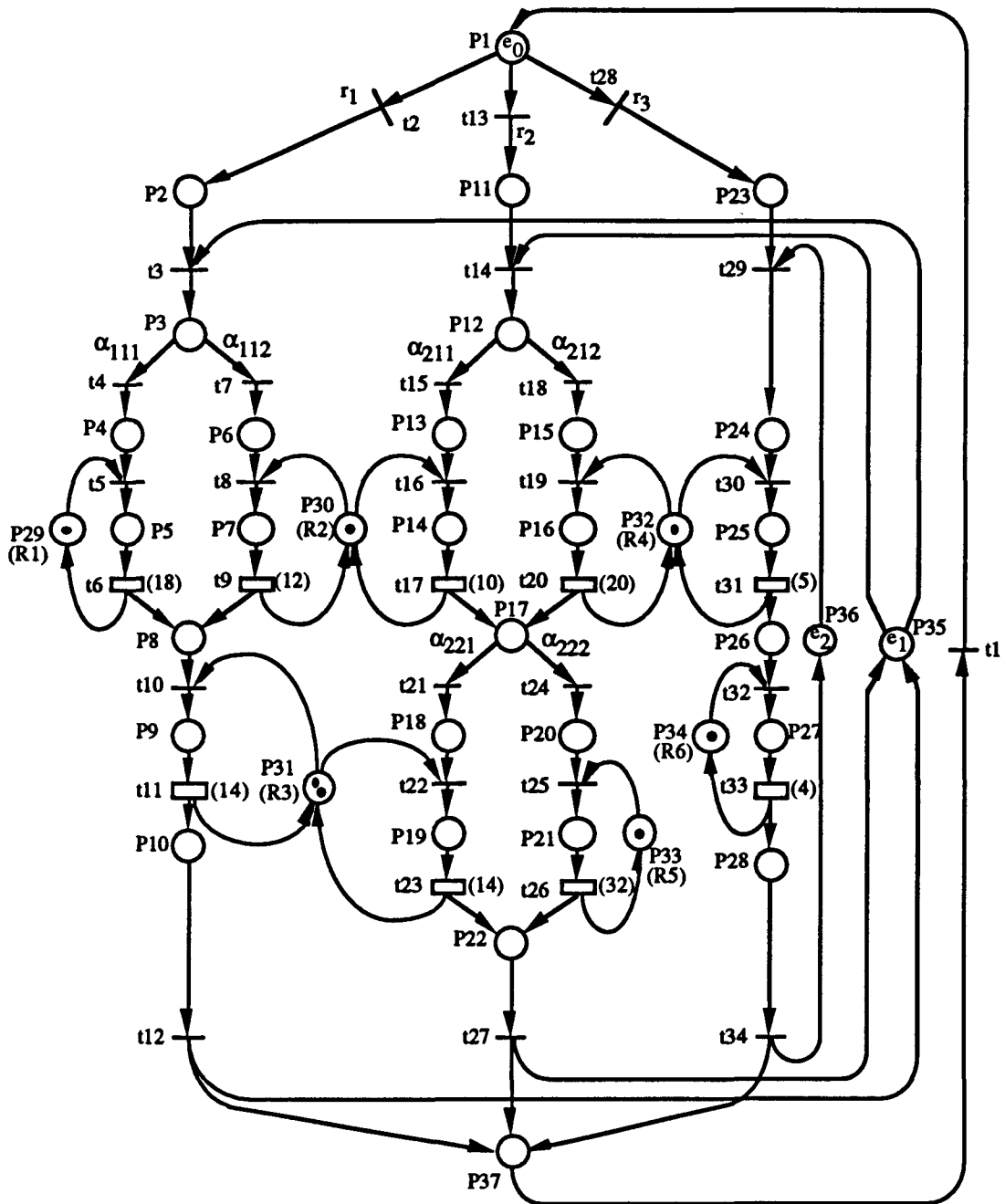


Fig. III-14 Modèle bouclé

Dans [CAM 90], les RdP ont été caractérisés en fonction de la manière de déterminer le vecteur des taux de visite. Il a été démontré que ce vecteur, noté $w^{(1)}$, s'il est normalisé par rapport à la transition t_1 , peut a priori être donné par

$$w^{(1)} = f(N, M_0, \alpha, \mu) \quad (\text{III.10})$$

où N est la structure du RdP, M_0 son marquage initial, α le vecteur des taux moyens de routage associés aux conflits (comprenant ici les conflits de routage et les ratios de production) et μ le vecteur contenant les durées associées aux transitions. Selon les classes des RdP, cette relation se simplifie. Par exemple pour un graphe d'événements avec un seul T-semiflot nous avons $w^{(1)} = f(N)$. Si nous effectuons un bouclage extérieur commun en modélisant le conflit au niveau de la place P_1 par un choix libre, exactement de la façon dont nous l'avons déjà introduit pour la modélisation des flexibilités, les taux de visite de toutes les transitions dépendent seulement de la structure du RdP et des taux de routage associés aux conflits.

Nous insistons ici sur le fait que notre modèle n'est pas un RdP à choix libre, car la résolution des conflits au niveau des ressources partagées est conditionnée par la présence ou l'absence de marques dans les gammes. Tous les taux de visite sont cependant définis de manière unique par des taux de routage associés à des choix libres. En respectant la terminologie de [CAM 90], notre modèle est un RdP à *T-semiflots librement reliés* (FRT-net, *freely related T-semiflow*).

Nous appelons C l'ensemble des conflits représentatifs des flexibilités et repérons chaque conflit par le couple d'indices (i,j) qui désigne le j -ème conflit de la i -ème gamme. A tout $c_{ij} \in C$ nous associons l'ensemble $\alpha_{ij} = \{\alpha_{ijk}\}_{k=1..K_{ij}}$ des taux de routage des K_{ij} choix du conflit. Nous supposons que

$$\begin{aligned} \alpha_{ijk} &\in IQ^+ & k &= 1..K_{ij} \\ \sum_k \alpha_{ijk} &= 1 & i &= 1..I \quad j = 1..J_i \end{aligned} \quad (\text{III.11})$$

Définition 6: Nous appelons *routage*, noté α , un ensemble de α_{ij} associés aux conflits $c_{ij} \in C$ et vérifiant les contraintes III.11.

III.3.4 Caractérisation structurelle

Dans le cas général, les invariants d'un RdP peuvent être déterminés avec un algorithme spécifique [MAR 82][TOU 82][ALA 82][ALA 89] mais comme conséquence de la construction systématique, nous pouvons directement énumérer les invariants de notre modèle. Il existe $R+H+1$ P-semiflots à supports minimaux, associés respectivement aux types de machines (Y_1, \dots, Y_R), aux bouclages intérieurs (Y_{R+1}, \dots, Y_{R+H}) et au bouclage extérieur (Y_{R+H+1}). Pour l'exemple de la figure III-14, nous avons $R=6$, $H=2$ et donc les P-semiflots suivants:

$$\begin{aligned}
 Y_1 &= (00001 \ 00000 \ 00000 \ 00000 \ 00000 \ 00010 \ 00000 \ 00)^T \\
 Y_2 &= (00000 \ 01000 \ 00010 \ 00000 \ 00000 \ 00001 \ 00000 \ 00)^T \\
 Y_3 &= (00000 \ 00010 \ 00000 \ 00010 \ 00000 \ 00000 \ 10000 \ 00)^T \\
 Y_4 &= (00000 \ 00000 \ 00000 \ 10000 \ 00001 \ 00000 \ 01000 \ 00)^T \\
 Y_5 &= (00000 \ 00000 \ 00000 \ 00000 \ 10000 \ 00000 \ 00100 \ 00)^T \\
 Y_6 &= (00000 \ 00000 \ 00000 \ 00000 \ 00000 \ 00100 \ 00010 \ 00)^T \\
 Y_7 &= (00111 \ 11111 \ 01111 \ 11111 \ 11000 \ 00000 \ 00001 \ 00)^T \\
 Y_8 &= (00000 \ 00000 \ 00000 \ 00000 \ 00011 \ 11110 \ 00000 \ 10)^T \\
 Y_9 &= (11111 \ 11111 \ 11111 \ 11111 \ 11111 \ 11100 \ 00000 \ 01)^T
 \end{aligned}$$

Puisque notre RdP est couvert par des P-semiflots, il est borné. De plus, il est ordinaire et les seules synchronisations existant dans le modèle concernent des P-semiflots distincts. Par conséquent, le nombre de marques dans chaque P-semiflot est constant. Nous pouvons représenter les couplages (synchronisations) entre les P-semiflots par un schéma comme celui de la figure ci-dessous.

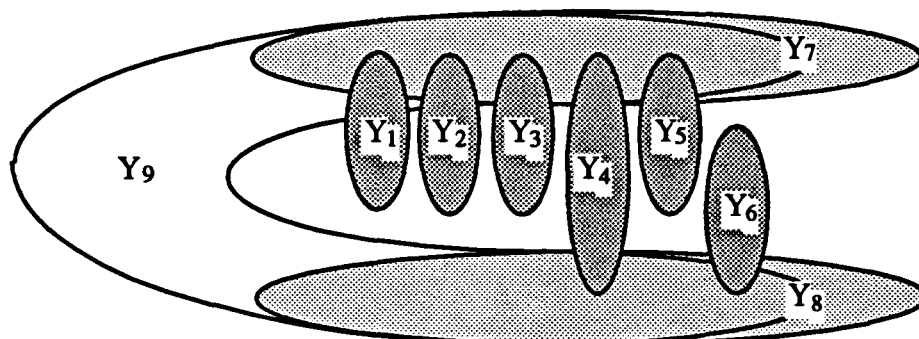


Fig. III-15 Couplages entre les P-semiflots

III.4 Evaluation préliminaire

III.4.1 Mise en équation

Dans le cas général, l'optimisation du fonctionnement de l'atelier concerne toutes les classes de décisions que nous avons identifiées auparavant. Il est cependant possible que certaines décisions soient imposées par un autre niveau, par exemple pour tenir compte de contraintes particulières. Nous considérons ici le cas où les types de produits et les ratios associés sont donnés. Cette situation peut se présenter par exemple lors d'une phase de planification fine où une certaine configuration de l'atelier doit être évaluée. Il est également possible qu'il s'agisse d'élaborer la commande du système en partant d'une planification donnée.

Nous supposons donc que les r_i sont donnés par (III.1) et que seul le routage α reste à déterminer. Pour cela, nous explicitons dans un premier temps les relations qui lient le débit de production et l'horizon à un routage α qui peut être a priori quelconque.

III.4.1.1 Expression du débit de production

Puisque nous supposons ici qu'un niveau d'en-cours suffisant permet de découpler les P-semiflots de notre modèle, le débit de production maximal est imposé par une ou plusieurs ressources critiques, appelées également ressources menantes. Une ressource menante se définit à partir de sa charge associée. Nous allons développer l'expression de cette charge en nous appuyant sur le modèle RdP développé précédemment. Nous définissons d'abord la charge associée à une transition:

Définition 7: La charge associée à une transition t_j est le produit du délai $\mu(t_j)$ associé à cette transition par son taux de visite, $w^{(1)}(t_j)$.

$$Z_t^{(1)}(t_j) = \mu(t_j) w^{(1)}(t_j) \quad t_j \in T \quad (\text{III.12})$$

Nous notons $Z_t^{(1)}$ le vecteur des charges de toutes les transitions du RdP, normalisé sur t_1 . Une composante j de ce vecteur peut être interprétée comme le temps moyen que la transition t_j impose entre deux tirs de la transition t_1 .

La prochaine étape consiste à relier la charge des transitions à celles des P-semiflots. En effet, en multipliant un P-semiflot Y par la matrice de pré-incidence (ou post-incidence car $Y^T (\text{POST} - \text{PRE}) = 0$) nous obtenons alors un vecteur de dimension égale au nombre de transitions du RdP et dont la j -ème composante est donnée par l'expression $\sum_{p \in |Y|} \text{pré}(p, t_j)$.

Un seul élément de cette somme est non nul, puisqu'un P-semiflot de notre RdP ne contient jamais deux places reliées à la même transition par une synchronisation. Puisque notre RdP est ordinaire, la j -ème composante du produit $Y^T \text{PRE}$ est alors égale à un si et seulement si la transition t_j se trouve située entre deux places du support de Y . Nous pouvons alors définir la charge associée à un P-semiflot ainsi que celle associée à chacune des marques:

Définition 8: La charge associée à un P-semiflot Y_i , notée $Z_Y^{(1)}(i)$ et normalisée sur la transition t_1 est donnée par l'expression

$$Z_Y^{(1)}(i) = Y_i^T \text{PRE} Z_t^{(1)} \quad i = 1..R+H+1 \quad (\text{III.13})$$

Puisque nous avons pu associer chaque type de ressource à un P-semiflot, la charge associée à une marque d'un P-semiflot correspond à la charge moyenne d'une ressource - en supposant une répartition parfaite de cette charge sur toutes les ressources de ce type.

Définition 9: La charge associée à une marque d'un P-semiflot Y_i , normalisée sur la transition t_1 , est définie par le rapport entre la charge associée au P-semiflot et le nombre constant et non nul de marques

$$Z_e^{(1)}(i) = \frac{Y_i^T \text{PRE} Z_t^{(1)}}{Y_i^T M_0} \quad i = 1..R+H+1 \quad (\text{III.14})$$

Par conséquent, la charge maximale qui va imposer le débit de l'atelier est donnée par l'expression

$$Z_{\max}^{(1)} = \max_{r=1..R} \left\{ \frac{Y_r^T \text{PRE} Z_t^{(1)}}{Y_r^T M_0} \right\} \quad (\text{III.15})$$

Remarquons que cette expression est équivalente à la *vitesse propre* associée à un P-semiflot définie dans [SIF 80] pour une temporisation des places. De même, dans [CAM 90] une expression similaire a été obtenue pour des classes de RdP plus générales, notamment pour une minoration du temps de cycle dans le cas de temporisations stochastiques.

La valeur de Z_{\max} dépend du routage α à travers le vecteur $Z_t^{(1)}$. Pour un routage α donné, le débit de production, mesuré en nombre de produits par unité de temps, est noté Φ_α et donné par

$$\Phi_\alpha = \frac{1}{\max_{r=1..R} \left\{ \frac{Y_r^T \text{PRE } Z_t^{(1)}}{Y_r^T M_0} \right\}} \quad (\text{III.16})$$

Il se pose alors la question de savoir sous quelles conditions un fonctionnement cyclique (CEF ou CCF) avec un débit Φ_α peut être réalisé.

III.4.1.2 Expression de l'horizon pour un fonctionnement cyclique

Nous supposons qu'un routage α est spécifié par ses coefficients α_{ijk} tels que

$$\alpha_{ijk} = \frac{p_{ijk}}{q_{ij}} \quad \text{avec } p_{ijk} \in \mathbb{N} \quad q_{ij} \in \mathbb{N}^* \quad (\text{III.17})$$

$$\text{pgcd}(p_{ijk}, q_{ij}) = 1 \quad i = 1..I \quad j = 1..J_i \quad k = 1..K_{ij}$$

Pour pouvoir effectivement réaliser un taux de routage α_{ijk} , il faut disposer exactement de q_{ij} , ou d'un multiple de q_{ij} , décisions discrètes. Afin de réaliser un fonctionnement cyclique pendant lequel f_i (CEF) ou un multiple de f_i (CCF) produits de type i sont fabriqués tout en respectant en même temps un partage de flux spécifié par un dénominateur q_{ij} , il faut fabriquer au minimum le *plus petit commun multiple* de f_i et q_{ij} , noté $\text{ppcm}(f_i, q_{ij})$.

Le plus petit horizon pour lequel cette condition est vérifiée est alors donné par

$$v_{ij} = \frac{\text{ppcm}(q_{ij}, f_i)}{f_i} \quad \forall (i, j) \text{ tq } c_{ij} \in C \quad (\text{III.18})$$

Puisque cette condition doit être vérifiée pour tous les conflits dans toutes les gammes du RdP, l'horizon minimal pour un fonctionnement cyclique est donné par

$$v = \text{ppcm}_{i=1..J} \left(\frac{\text{ppcm}(f_i, \{q_{ij}\}_{j=1..J_i})}{f_i} \right) \quad (\text{III.19})$$

La relation III.19 associe donc à des ratios et un routage α donnés, l'horizon minimal sur lequel une réalisation discrète peut être espérée. Dans la suite, nous appelons v_α cet horizon.

III.4.1.3 Prise en compte de la nature discrète des opérations

La relation III.19 n'exprime qu'une condition nécessaire pour la réalisation effective de Φ_α par un fonctionnement cyclique. En effet, la charge d'une ressource individuelle a été obtenue en divisant simplement la charge associée à un type de ressource par le nombre de ressources de ce type (cf. III.14). Une telle répartition optimale ne peut pas être espérée dans le cas général.

Afin de répartir la charge sur les ressources individuelles d'un même type, il faut détailler la composition de cette charge. Si nous supposons que v_α est dans un premier temps donné par III.19, nous pouvons déterminer le nombre de fois qu'une transition $t_j \in T$ sera tirée pendant un fonctionnement cyclique d'horizon v_α .

Nous notons $\beta(t_j)$ ce nombre qui est donné par le produit du taux de visite $w^{(1)}(t_j)$ par le nombre de tirs de t_1 pendant un fonctionnement cyclique d'horizon v_α :

$$\beta(t_j) = w^{(1)}(t_j) v_\alpha \sum_i f_i \quad \forall t_j \in T \quad (\text{III.20})$$

$$\beta(t_j) \in \mathbb{N}$$

Les $\beta(t_j)$ sont des entiers dont le plus grand commun diviseur est égal à 1

$$\text{pgcd}\left(\{\beta(t_j)\}_{t_j \in T}\right) = 1 \quad (\text{III.21})$$

En effet, supposons qu'il existe un diviseur commun $\delta > 1$. Alors v_α pourrait être divisé par ce nombre et les ratios et le routage α pourraient être respectés sur un horizon plus court que v_α . Ceci est impossible car v_α est par construction le plus petit horizon respectant les ratios et le routage α .

A partir des $\beta(t_j)$ nous pouvons définir l'ensemble des opérations associées à chaque type de ressource pour un fonctionnement cyclique d'horizon v_α .

Définition 10: L'ensemble des opérations associées à un type de ressource r pour un fonctionnement cyclique d'horizon v_α , noté $B_{v_\alpha}(r)$, contient exactement $\beta(t_j)$ fois la transition t_j si et seulement si t_j est associée au type de ressource r .

Répartir les opérations associées à un type de ressources sur des ressources individuelles, c'est à dire résoudre les indéterminismes associés à la classe D4, consiste alors à déterminer pour chaque type $r = 1..R$ une partition \mathcal{P}_r de $B_{v_\alpha}(r)$ en m_r sous-ensembles $B_{v_\alpha}(r, m)$, $m = 1..m_r$, tels que

$$\bigcup_{m=1..m_r} B_{v_\alpha}(r, m) = B_{v_\alpha}(r) \quad r = 1..R \quad (\text{III.22})$$

$$\bigcap_{m=1..m_r} B_{v_\alpha}(r, m) = \emptyset \quad r = 1..R \quad (\text{III.23})$$

Cette partition de l'ensemble des opérations associées à un type de ressources permet alors de définir la charge de chaque ressource individuelle $m = 1..m_r$ de type r , notée $Z_e^{v_\alpha}(r, m)$

$$Z_e^{v_\alpha}(r, m) = \sum_{t \in B_{v_\alpha}(r, m)} \mu(t) \quad m = 1..m_r \quad r = 1..R \quad (\text{III.24})$$

Remarquons que l'expression III.24 de la charge d'une machine individuelle est relative à la fabrication de $v_\alpha \sum f_i$ produits pour un fonctionnement cyclique d'horizon v_α , contrairement à l'expression $Z_e^{(1)}$ de III.14 donnant la charge d'une machine 'moyenne' pour la fabrication d'un seul produit.

Généralement il existe plusieurs façons de partitionner la charge associée à un type de ressources. Dans le cas d'un équilibrage parfait entre les m_r exemplaires d'un même type nous avons

$$Z_e^{v_\alpha}(r, m) = (v_\alpha \sum_i f_i) Z_e^{(1)}(r) \quad m = 1..m_r \quad r = 1..R \quad (\text{III.25})$$

Malheureusement, cette répartition parfaite n'est pas toujours possible. En effet, si les charges de tous les types de ressources dont il existe plusieurs exemplaires ont été réparties sur les ressources individuelles, il est possible que la nouvelle charge maximale soit plus grande qu'avant. De plus, cette nouvelle charge peut être associée à une ou plusieurs ressources qui n'étaient pas critiques précédemment.

Nous pouvons alors formuler une deuxième condition pour qu'un débit Φ_α puisse être atteint sur un horizon v_α :

$\forall r = 1..R$, tq $m_r \geq 2$, $\exists \mathcal{P}_r$ de $B_{v_\alpha}(r)$, respectant III.22 et III.23 et telle que

$$\max_{m = 1..m_r} \left\{ Z_e^{v_\alpha}(r, m) = \sum_{t \in B_{v_\alpha}(r, m)} \mu(t) \right\} \leq \frac{v_\alpha \sum_i f_i}{\Phi_\alpha} \quad (\text{III.26})$$

Montrons que les conditions III.19 et III.26, associées à l'hypothèse d'un en-cours suffisant, sont effectivement nécessaires et suffisantes pour la réalisation d'un débit de production Φ_α sur un horizon v_α :

◆ De l'analyse effectuée précédemment, il découle que les deux conditions sont nécessaires. En effet, rappelons que v_α est par construction le plus petit horizon sur lequel un routage α et les ratios r_i peuvent être réalisés par des décisions discrètes. La condition III.19 est donc nécessaire. Rappelons encore que nous ne considérons pas les

problèmes cumulatifs dans lesquels x ressources peuvent effectuer une opération x fois plus vite. Chaque opération doit donc être associée à une et une seule ressource. La répartition des opérations associées à un type de ressource est donc nécessaire et par conséquent la condition III.26 l'est aussi.

◆ Inversement, si la condition III.19 est vérifiée, il est possible d'extraire des gammes un ensemble de chemins (suites déterministes d'opérations correspondant aux T-semiflots) tels que le routage α et les ratios soient respectés. Si la condition III.26 est vérifiée, il est possible d'affecter chaque opération à une ressource individuelle, devant effectuer cette opération une et une seule fois par cycle. Si nous plaçons une pièce semi-finie pour chaque opération dans le système, de sorte que chaque ressource puisse effectuer les opérations qui lui sont affectées sans délai, alors le système fonctionne avec le débit Φ_α . □

Un autre résultat découle de cette démonstration: un majorant de l'en-cours nécessaire pour réaliser un débit Φ_α avec un fonctionnement cyclique est donné par le nombre d'opérations à effectuer pendant un cycle (élémentaire ou complexe). Remarquons que ce majorant est généralement très important, surtout lorsque v_α est grand. Nous reviendrons sur le problème de la taille de l'en-cours dans le chapitre IV.

Il est évidemment possible qu'il n'existe pas de partitions des ensembles $B_{v_\alpha}(r)$ telles que la condition III.26 soit vérifiée pour tous les types de ressources. Par conséquent, le débit Φ_α ne peut être atteint sur un horizon v_α .

Deux possibilités apparaissent pour atteindre malgré tout un débit Φ_α :

◆ Premièrement, il est possible que le même débit puisse être réalisé avec un autre routage α' . En effet, certaines des composantes d'un routage α peuvent éventuellement varier sans que le débit global ne soit affecté. Cette insensibilité du débit vis-à-vis d'une variation de certaines paramètres provient de l'opérateur *max* dans l'expression III.16. Un routage $\alpha' \neq \alpha$ peut être associé au même horizon v_α , il engendre toujours des $\beta'(t_j)$ différents des $\beta(t_j)$, pour au moins quelques $t_j \in T$ (cf. III.20, définition de $\beta(t_j)$). Les transitions apparaissent alors avec des taux de visite différents et les charges des différents types de ressources sont modifiées. D'autres partitions des

charges devenant possibles et il se peut que la condition III.26 soit vérifiée pour tous les types de ressources.

◆ Une deuxième possibilité consiste à garder le même routage α et à augmenter l'horizon en considérant des multiples de v_α . Nous posons donc

$$v_{\zeta\alpha} = \zeta v_\alpha \quad \zeta \in \mathbb{N}^* \quad (\text{III.27})$$

$$B_{v_\alpha}^\zeta(r) = \bigcup_{1.. \zeta} B_{v_\alpha}(r) \quad r = 1..R$$

où l'ensemble $B_{v_\alpha}^\zeta(r)$ contient ζ fois les éléments de $B_{v_\alpha}(r)$. Le nombre d'opérations à répartir croît alors linéairement avec ζ , de nouvelles partitions deviennent ainsi possibles.

Pour un type de ressources $r = 1..R$ donné, une partition garantissant un équilibrage parfait entre les ressources individuelles de ce type existe, si ζ est un multiple de m_r , $\zeta = km_r$, $k \in \mathbb{N}^*$. En effet, dans ce cas $B_{v_\alpha}^\zeta(r)$ peut être scindé en m_r sous-ensembles identiques $B_{v_\alpha}^k(r, m)$, $m = 1..m_r$, tels que

$$B_{v_\alpha}^\zeta(r) = \bigcup_{m=1..m_r} B_{v_\alpha}^k(r, m) \quad (\text{III.28})$$

Si nous appelons Π l'ensemble des indices des types de ressource qui ne vérifient pas la condition III.26, une condition pour que Φ_α soit réalisable sur un horizon $v_{\zeta\alpha}$ est donc donnée par

$$\zeta = \text{ppcm}_{r \in \Pi}(m_r) \quad (\text{III.29})$$

Cette condition ne concerne que les types de ressources $r \in \Pi$. Pour les autres types de ressources $r \in \{1..R\} \setminus \Pi$, pour lesquels une partition respectant le débit Φ_α sur l'horizon v_α existait, nous observons une linéarité en ζ des deux côtés de l'inégalité de III.26. En effet, à droite de l'inégalité III.26 nous avons $v_{\zeta\alpha} = \zeta v_\alpha$. A gauche de

l'inégalité nous introduisons pour tout $m = 1..m_r$ l'ensemble $B_{v_\alpha}^\zeta(r, m)$ contenant ζ fois l'ensemble $B_{v_\alpha}(r, m)$. Il vient

$$\sum_{t \in B_{v_\alpha}^\zeta(r, m)} \mu(t) = \zeta \left(\sum_{t \in B_{v_\alpha}(r, m)} \mu(t) \right) \quad r \in \{1..R\} \setminus \Pi \quad (\text{III.30})$$

A cause de cette linéarité des deux côtés de l'inégalité de III.26, la condition III.26 est vérifiée pour tout multiple de l'horizon v_α . Par conséquent, III.29 est une condition suffisante pour que le débit Φ_α puisse être réalisé sur l'horizon $v_{\zeta\alpha}$.

Nous pouvons encore améliorer cette condition en tenant compte des valeurs numériques des $\beta(t_j)$. Désignons par $r = \text{aff}(t_j)$ le type de ressource associé à une transition t_j et calculons δ donné par

$$\delta = \text{pgcd} \left(m_r, \{ \beta(t_j) \}_{\text{aff}(t_j)=r} \right) \quad r \in \Pi \quad (\text{III.31})$$

Malgré la propriété III.21, qui concerne toutes les transitions du RdP, il est possible que $\delta > 1$, car III.31 concerne seulement le nombre m_r et les $\beta(t_j)$ d'un sous-ensemble de transitions de T .

Si $\delta > 1$, alors il existe une partition de la charge associée au type de ressource r en m_r sous-ensembles identiques pour $\zeta = k \frac{m_r}{\delta}$, $k \in \mathbb{N}^*$. En tenant compte de tous les types de ressources, il suffit donc de prendre ζ défini par

$$\zeta = \text{ppcm}_{r \in \Pi} \left(\frac{m_r}{\text{pgcd} \left(m_r, \{ \beta(t_j) \}_{\text{aff}(t_j)=r} \right)} \right) \quad (\text{III.32})$$

Nous pourrions encore améliorer cette condition en tenant compte des durées associées aux transitions. En effet, si deux transitions distinctes mais associées au même type de ressource ont la même durée, $t_i \neq t_j$ telles que $\mu(t_i) = \mu(t_j)$ et $\text{aff}(t_i) = \text{aff}(t_j)$, elles

pourraient être comptabilisées ensemble dans le calcul du pgcd de l'expression III.32. Une plus petite valeur de ζ pourrait en résulter à cause de l'inégalité

$$\text{pgcd}(a+b, c) \geq \text{pgcd}(a, b, c) \quad \forall (a, b, c) \in \mathbb{N}^3 \quad (\text{III.33})$$

Nous ne poursuivons pas l'analyse dans cette direction mais nous pouvons remarquer que la considération sur le regroupement des transitions de même durée permet d'affirmer que les conditions III.29 et III.32 ne sont pas nécessaires mais seulement suffisantes.

III.4.2 Détermination du routage

III.4.2.1 Formulation du problème

Dans la section précédente nous avons vu que les expressions du débit et de l'horizon sont des fonctions non linéaires des éléments d'un routage α . Il est donc difficile d'envisager une inversion de ces fonctions contenant des opérateurs tels que *max*, *ppcm* ou *pgcd*. Par conséquent, l'analyse du couple débit-horizon doit s'appuyer sur une représentation paramétrée en α

$$\begin{aligned} \Phi_\alpha &= \Phi(\alpha) \\ v_\alpha &= v(\alpha) \end{aligned} \quad (\text{III.34})$$

Une optimisation conjointe du débit et de l'horizon pourrait être effectuée en introduisant une fonction de coût, composant ces deux indices selon l'expression linéaire suivante

$$g(\alpha) = g_v \frac{v(\alpha)}{v_{\text{ref}}} + g_\Phi \frac{\Phi(\alpha)}{\Phi_{\text{ref}}} \quad (\text{III.35})$$

Dans cette fonction nous avons séparé les coefficients de pondération, g_v et g_Φ , des coefficients de normalisation, v_{ref} et Φ_{ref} . Ces derniers servent de référence afin de rendre homogènes des indices de nature très différente. Les problèmes liés à une telle approche se situent cependant à plusieurs niveaux:

◆ Tout d'abord, dans la section précédente nous avons vu que la réalisation effective d'un fonctionnement cyclique avec Φ_α et v_α obtenus par III.16 et III.19 ne peut pas être garantie s'il existe plusieurs ressources du même type. En effet, la répartition des opérations sur des ressources individuelles peut révéler des incompatibilités avec Φ_α et v_α et donc entraîner des performances réelles réduites, caractérisées par une diminution du débit, $\Phi'_\alpha < \Phi_\alpha$, ou une augmentation de l'horizon, $v'_\alpha > v_\alpha$.

◆ Un deuxième type de problèmes concerne la sensibilité différente des deux indices vis-à-vis d'une variation de la valeur d'un élément de α . Bien que le débit dépende des éléments de α à travers l'inverse de l'opérateur *max*, la dépendance est continue. Une petite variation au niveau du routage entraînera une petite - ou éventuellement aucune - variation du débit. En revanche, l'horizon peut varier de façon discontinue si la valeur d'un élément de α passe par exemple de $\alpha_{ijk} = \frac{49}{100}$ à $\alpha'_{ijk} = \frac{1}{2}$.

Ces remarques indiquent qu'une formulation de type III.35 n'est pas suffisante pour l'optimisation conjointe des deux indices et qu'il faut procéder à une analyse plus détaillée. Dans la suite nous considérons deux approches plus particulières du problème, la première mettant en avant le débit, la deuxième l'horizon de réalisation effective.

III.4.2.2 Maximisation du débit de production

Une formulation du problème de maximisation du débit est donnée par

$$\left| \begin{array}{l} \text{minimiser} \left(\max_{r=1..R} \left\{ \frac{Y_r^T \text{PRE } Z_t^{(1)}}{Y_r^T M_0} \right\} \right) \\ \alpha_{ijk} \geq 0 \quad k = 1..K_{ij} \\ \sum_k \alpha_{ijk} = 1 \quad i = 1..I, j = 1..J_i \end{array} \right. \quad (\text{III.36})$$

Dans [FRE 88] cette expression non-linéaire est transformée en posant

$$x = \max_{r=1..R} \left\{ \frac{Y_r^T \text{PRE } Z_t^{(1)}}{Y_r^T M_0} \right\} \quad (\text{III.37})$$

Le problème se présente alors sous forme d'un problème de programmation linéaire dans lequel x et les α_{ijk} sont les variables à déterminer

$$\left. \begin{array}{l} \text{minimiser } x \\ x \geq \frac{Y_r^T \text{PRE } Z_t^{(1)}}{Y_r^T M_0} \quad r = 1..R \\ \alpha_{ijk} \geq 0 \quad k = 1..K_{ij} \\ \sum_k \alpha_{ijk} = 1 \quad i = 1..I, j = 1..J_i \end{array} \right\} \quad (\text{III.38})$$

Par l'utilisation de l'algorithme du Simplexe ce problème peut généralement être résolu rapidement*. Les résultats de cette étape de résolution sont la valeur de x , les indices des types de ressources menantes et les valeurs d'une partie, éventuellement vide, des taux de routage.

Nous désignons par α^1 le sous-ensemble des éléments de α dont les taux ont été fixés et par α^2 le complément, tels que

$$\begin{aligned} \alpha^1 \cap \alpha^2 &= \emptyset \\ \alpha^1 \cup \alpha^2 &= \{ \alpha_{ijk} \}_{i=1..I, j=1..J_i, k=1..K_{ij}} \end{aligned} \quad (\text{III.39})$$

Remarquons que α^1 ou α^2 peuvent être vides. Dans le premier cas, aucun taux de routage n'a été fixé par l'optimisation, ce qui signifie que les ressources critiques ne peuvent pas être déchargées en orientant les flux de produits sur d'autres types de ressources. Dans le deuxième cas tous les paramètres ont été fixés et nous sommes dans les conditions d'application des résultats de la section précédente sur la réalisation d'un routage α donné.

* bien que l'algorithme du simplexe soit très performant en pratique, il a une complexité non-polynomiale dans des cas défavorables. Un algorithme polynomial peut être trouvé dans [KAR 84].

Dans le cas général, il faut supposer que seule une partie des taux de routages a été fixée par la programmation linéaire. En reportant les valeurs des éléments de α^1 ainsi que x dans les contraintes de III.36, nous obtenons l'ensemble des contraintes que les taux de routage non encore fixés doivent vérifier. Ces contraintes sont notées C_{α^2} .

Nous appelons horizon minimal, noté v_{α^1} , l'horizon imposé par les seuls taux de routage de l'ensemble α^1

$$v_{\alpha^1} = \text{ppcm}_{i=1..I} \left(\frac{\text{ppcm}_{(i,j) \text{ si } \exists k \in 1..K_{ij} \text{ tq } \alpha_{ijk} \in \alpha^1} \left(f_i, \{q_{ij}\} \right)}{f_i} \right) \quad (\text{III.40})$$

Le débit maximum, donné par l'inverse de x , est par conséquent noté

$$\Phi_{\alpha^1} = \frac{1}{x} \quad (\text{III.41})$$

Il reste à fixer les taux de routage appartenant à l'ensemble α^2 . Dans [FRE 88] une *optimisation secondaire* est proposée consistant à utiliser les degrés de liberté des éléments de α^2 pour équilibrer au mieux les charges des ressources non critiques. Un tel équilibrage se justifie parce qu'il permet d'écartier les charges des ressources qui ne sont pas en situation de criticité de celles des ressources critiques. Dans un environnement perturbé, par exemple, cette précaution peut aider à éviter des changements inopinés des ressources critiques (*shifting bottleneck*), difficile à prendre en compte dans un contexte de pilotage temps réel.

Puisque nous considérons par hypothèse un régime de fonctionnement stable, nous n'effectuons pas d'optimisation secondaire, mais gardons tous les degrés de liberté afin de limiter le plus possible l'horizon sur lequel Φ_{α^1} peut être réalisé.

Pour $\lambda \in \mathbb{N}^*$ nous posons

$$v_{\lambda\alpha^1} = \lambda v_{\alpha^1} \quad \lambda \in \mathbb{N}^* \quad (\text{III.42})$$

Pour tout $\lambda \in \mathbb{N}^*$, un ensemble d'éléments de α^2 qui ne prolonge pas l'horizon $v_{\lambda\alpha^1}$ est noté $\alpha_2^{v_{\lambda\alpha^1}}$. Un tel ensemble est défini par

$$\alpha_2^{v_{\lambda\alpha^1}} = \left\{ \alpha_{ijk} \mid \alpha_{ijk} \in \alpha^2, \alpha_{ijk} = \frac{p_{ijk}}{v_{\lambda\alpha^1} f_i}, p_{ijk} \in \mathbb{N}, \sum_k \alpha_{ijk} = 1 \right\} \quad (\text{III.43})$$

L'ensemble des routages $\alpha_2^{v_{\lambda\alpha^1}}$ définis par III.43 et respectant en même temps les contraintes C_{α^2} est noté $A_2^{v_{\lambda\alpha^1}}$. Si cet ensemble est vide pour $\lambda = 1$, le débit maximal ne peut pas être atteint sur v_{α^1} et il faut itérer en posant successivement $\lambda = 2, 3, \dots$. Avec un dénominateur croissant, les α_{ijk} sont de plus en plus denses et il existe donc un λ_0 , tel que $A_2^{v_{\lambda_0\alpha^1}}$ est non vide.

S'il existe un élément $\alpha \in A_2^{v_{\lambda_0\alpha^1}}$ pour lequel la condition III.26, concernant la répartition des charges sur des ressources individuelles, est vérifiée, le débit peut être atteint sur l'horizon $v_{\lambda_0\alpha^1}$.

Sinon, il faut prolonger l'horizon, ce qui peut être effectué de deux manières différentes. Une première consiste à itérer à partir de III.42 et III.43 avec un coefficient $\lambda > \lambda_0$. Alternativement, il est possible de garder le coefficient λ_0 et un routage $\alpha \in A_2^{v_{\lambda_0\alpha^1}}$ et d'augmenter l'horizon en considérant des multiples de $v_{\lambda_0\alpha^1}$ comme proposé dans la présentation des conditions III.29 et III.32.

Nous pouvons par conséquent affirmer que le débit maximal, obtenu avec un équilibrage de flux peut effectivement être atteint par un fonctionnement cyclique. Notons cependant que cet équilibrage de flux ne tient pas compte de l'horizon nécessaire pour une réalisation effective. En effet, pour des cas réels où les taux de routage sont a priori quelconques, la valeur de v_{α^1} peut devenir trop importante. Il apparaît donc que la maximisation du débit ne peut pas être totalement découplée de l'horizon et qu'il faut essayer de maîtriser simultanément ces deux indices de performance.

Une possibilité de limiter la valeur de v_{α^1} consiste à approximer les valeurs numériques des taux de routage fixées par l'équilibrage des ressources par des fractions

avec des dénominateurs 'raisonnables'. Puisque le débit est une fonction continue des taux de routages, nous pouvons en attendre une diminution de débit pas trop importante lors d'une faible variation de ces taux.

III.4.2.3 Minimisation de l'horizon

Au lieu de considérer le problème avec $v = 1$, nous nous plaçons dans le cas plus général avec $v = v_0$. Cette formulation est plus restrictive que $v \leq v_{\max}$ et ne correspond donc pas au cas où l'horizon est pris en compte sous forme d'une contrainte. Cependant, en énumérant toutes les valeurs $v_0 = 1, 2, \dots, v_{\max}$, il est théoriquement possible de traiter le cas $v \leq v_{\max}$.

Puisque nous cherchons à réaliser un fonctionnement cyclique pendant lequel nous fabriquons exactement $v_0 f_i$ produits de type i , les taux de routage associés aux conflits de chaque gamme $i = 1..I$ doivent s'exprimer avec des dénominateurs compatibles avec le produit $v_0 f_i$. Un routage admissible pour l'horizon v_0 est alors donné par

$$\alpha^{v_0} = \left\{ \alpha_{ijk} \mid \alpha_{ijk} = \frac{p_{ijk}}{v_0 f_i}, p_{ijk} \in \mathbb{N}, \sum_k \alpha_{ijk} = 1, i = 1..I, j = 1..J_i, k = 1..K_{ij} \right\} \quad (\text{III.44})$$

Soit A^{v_0} l'ensemble de tous les routages admissibles pour l'horizon v_0 . Le débit maximal est alors donné pour un routage $\alpha_0 \in A^{v_0}$, non nécessairement unique, qui vérifie la relation III.45 dans laquelle $\Phi'(\alpha)$ est le débit obtenu après la répartition des charges sur des ressources individuelles.

$$\Phi_{\alpha_0}^{v_0} = \max_{\alpha \in A^{v_0}} \{ \Phi'(\alpha) \} \quad (\text{III.45})$$

L'ensemble A^{v_0} est fini et il est donc théoriquement possible de déterminer le meilleur routage. En pratique, lors d'une analyse de systèmes avec plusieurs flexibilités, nous sommes par contre confronté à un phénomène d'explosion combinatoire du nombre de solutions envisageables.

A cette complexité combinatoire s'ajoute le fait qu'il faut résoudre le problème de la répartition des charges sur des ressources individuelles pour chaque cas afin de connaître le débit $\Phi'(\alpha)$ effectivement réalisable. Puisque le parcours de l'espace des solutions est gourmand en temps de calcul, il peut être intéressant de déterminer une borne supérieure du débit avant d'entamer le parcours. Une telle borne est donnée par la valeur de $\Phi_{\alpha 1}$, explicitée dans la section précédente. Pendant le parcours de l'espace des solutions, chaque routage peut alors être comparé à cette borne supérieure et le parcours peut éventuellement être arrêté si une condition d'arrêt - un débit atteignant 95% de $\Phi_{\alpha 1}$ par exemple - est vérifiée.

III.4.3 Exemple d'application

Nous considérons le RdP de la figure III-16. Ce RdP correspond à l'exemple introduit précédemment (cf. figure III-9) et pour lequel nous avons explicité les P-semiflotts, T-semiflotts ainsi que le vecteur de taux de visite (cf. III.3.4). Les nombres entre parenthèses indiquent les temporisations que nous associons aux transitions représentatives des opérations.

Le RdP modélise un système avec trois produits et six types de ressources dont un type (R3) avec deux exemplaires identiques. Dans ce modèle apparaissent trois conflits dus à des flexibilités d'affectation, un conflit dans la gamme du premier type de produits (place P3) et deux conflits enchaînés dans la gamme du deuxième type de produits (places P12 & P17). L'en-cours est limité par deux contraintes, la première concerne les types de produits 1 et 2 (place P35), la deuxième contrainte le type de produits 3 (place P36). Nous supposons dans un premier temps que les quantités de marques e_0 , e_1 et e_2 sont suffisantes pour découpler les différents types de ressources.

Les trois types de produits doivent être fabriqués selon des ratios de production $r_1 = 0.4$, $r_2 = 0.2$ et $r_3 = 0.4$. Par conséquent, nous avons $f_1 = 2$, $f_2 = 1$ et $f_3 = 2$.

Compte tenu des P-semiflotts associés aux ressources et des taux de visite des transitions définis en III.3.4, les charges associées aux six types de ressources, données par la relation III.14, peuvent être explicitées.

Pour les six types de ressources nous obtenons alors les expressions suivantes

$$\left| \begin{array}{ll} Z_e^{(1)}(1) = \frac{7.2 \alpha_{111}}{1} & Z_e^{(1)}(4) = \frac{4 \alpha_{212} + 2}{1} \\ Z_e^{(1)}(2) = \frac{4.8 \alpha_{112} + 2 \alpha_{211}}{1} & Z_e^{(1)}(5) = \frac{6.4 \alpha_{222}}{1} \\ Z_e^{(1)}(3) = \frac{5.6 + 2.8 \alpha_{221}}{2} & Z_e^{(1)}(6) = \frac{1.6}{1} \end{array} \right. \quad (\text{III.47})$$

Dans la suite nous allons considérer les deux cas particuliers qui consistent respectivement à maximiser le débit et à déterminer l'horizon nécessaire, ou encore à minimiser l'horizon et à déterminer le plus grand débit atteignable.

III.4.3.1 Maximisation du débit

Afin de nous ramener à un problème linéaire, nous utilisons la transformation explicitée en III.37 et posons

$$x = \max_{r=1..6} \{Z_e^{(1)}(r)\} \quad (\text{III.48})$$

Nous pouvons ensuite résoudre le problème linéaire dans lequel x ainsi que les taux de routage sont des variables. La résolution indique que les types de ressources R1, R2 et R4 sont critiques. Les valeurs suivantes sont obtenues

$$\left| \begin{array}{l} x = 3.6 \\ \alpha^1 = \{\alpha_{111}, \alpha_{112}, \alpha_{211}, \alpha_{212}\} \text{ avec } \alpha_{111} = \alpha_{112} = 0.5 \\ \alpha_{211} = 0.6 \text{ et } \alpha_{212} = 0.4 \\ \alpha^2 = \{\alpha_{221}, \alpha_{222}\} \end{array} \right. \quad (\text{III.49})$$

Le débit maximal est donné par l'inverse de la plus grande charge (cf. III.16) et s'exprime selon la relation suivante

$$\Phi_{\alpha^1} = \frac{1}{x} = 0.278 \text{ pièces/u.t.} \quad (\text{III.50})$$

Calculons l'horizon minimal v_{α^1} sur lequel ce débit Φ_{α^1} est réalisable. Les valeurs numériques de α_{111} et α_{112} du conflit c_{11} s'expriment avec un dénominateur $q_{11}=2$, tandis que α_{211} et α_{212} du conflit c_{21} nécessitent un dénominateur $q_{21}=5$. En tenant compte des valeurs $f_1 = 2$, $f_2 = 1$ et $f_3 = 2$ et en utilisant l'expression III.40 nous obtenons

$$v_{\alpha^1} = \text{ppcm} \left(\frac{\text{ppcm}(2,2)}{2}, \frac{\text{ppcm}(1,5)}{1}, \frac{\text{ppcm}(2)}{2} \right) = 5 \quad (\text{III.51})$$

Nous pouvons remarquer que nous obtenons pour cet exemple une petite valeur de v_{α^1} en vertu des valeurs 'raisonnables' des taux de routage. Un fonctionnement répétitif respectant en même temps les ratios et les valeurs numériques des taux de visite spécifiées par l'optimisation peut donc être espéré pour un horizon $v_{\alpha^1} = 5$.

- Prise en compte des contraintes sur α^2 :

Encore faut-il déterminer les taux de routage non encore fixés par l'équilibrage des charges. En reportant la valeur de x ainsi que celles des éléments de α^1 dans les contraintes de la formulation III.38, nous obtenons les contraintes C_{α^2} suivantes

$$\frac{7}{16} \leq \alpha_{221} \leq \frac{4}{7} \quad \text{et} \quad \alpha_{221} + \alpha_{222} = 1 \quad (\text{III.52})$$

Pour que le débit Φ_{α^1} soit effectivement réalisable sur l'horizon $v_{\alpha^1}=5$, il faut qu'il existe des taux de routage α_{221} et α_{222} satisfaisant les contraintes C_{α^2} explicitées en III.52, tout en s'exprimant avec un dénominateur $q_{22}=f_2 v_{\alpha^1}=5$. Nous posons $\lambda=1$ pour déterminer l'ensemble des éléments vérifiant ces contraintes, noté $A_2^{v_{\alpha^1}}$ et spécifié en III.43.

En remarquant que $\frac{7}{16} = 0.4375$ et $\frac{4}{7} = 0.5714$, il est clair que nous ne pouvons pas déterminer de valeurs de α_{221} et α_{222} satisfaisant les contraintes III.52 avec un dénominateur $q_{22}=5$. Un horizon de $v_{\alpha^1}=5$ n'est donc pas suffisant. Il faut alors itérer avec des horizons $v_{\lambda\alpha^1} = \lambda v_{\alpha^1}$, $\lambda > 1$.

En posant $\lambda=2$ nous trouvons un seul ensemble de valeurs des taux de routage qui convienne

$$\alpha_2^{v_{\lambda\alpha^1}} = \left\{ \alpha_{221} = \frac{5}{10}, \alpha_{222} = \frac{5}{10} \right\} \quad (\text{III.53})$$

Par conséquent, la réalisation effective du débit Φ_{α^1} nécessite un horizon qui est au moins égal à $v_{\lambda\alpha^1}=10$. Le temps de cycle d'un CCF d'horizon $v_{\lambda\alpha^1}=10$ et de débit Φ_{α^1} est donné par la relation III.5

$$C^{10} = \frac{v \sum f_i}{\Phi_{\alpha^1}} = 180 \text{ u. t.} \quad (\text{III.54})$$

- Prise en compte de la partition des charges :

Il faut alors tester si la condition III.26 est vérifiée pour tous les types de ressources avec plusieurs exemplaires identiques. Le type de ressources R3 comporte deux exemplaires et il faut répartir la charge qui est associée à ce type de ressource pendant un CCF d'horizon $v_{\lambda\alpha^1}=10$.

En utilisant la relation III.20, nous trouvons $\beta(t_{11})=20$ et $\beta(t_{23})=5$. L'ensemble $B_{10}(3)$ associé aux deux ressources de type R3 contient alors 20 fois la transition t_{11} et 5 fois la transition t_{23} . Les transitions t_{11} et t_{23} ont les mêmes durées $\mu(t_{11}) = \mu(t_{23}) = 14$ u.t., il s'agit par conséquent de répartir 25 opérations de même durée sur deux ressources identiques.

La partition la plus régulière consiste à affecter 13 opérations à une ressource et 12 à l'autre, ce qui induit une charge de $13 \cdot 14$ u.t. = 182 u.t. à la ressource la plus chargée. Cette valeur est supérieure au C^{10} calculé en III.54 et le débit maximal ne pourra pas être atteint sur cet horizon. Par conséquent il faut encore allonger l'horizon.

◆ Une première possibilité consiste à redéfinir les taux de routage α_{221} et α_{222} avec un autre dénominateur. En posant $\lambda=3$, l'horizon passe à $v_{\lambda\alpha}=15$ et il faut donc essayer d'exprimer les taux de routage avec un dénominateur $q_{22}=f_2 v_{\lambda\alpha}=15$. Cette fois-ci deux solutions existent

$$A_2^{v_{\lambda\alpha}} = \left\{ \left\{ \alpha_{221} = \frac{7}{15}, \alpha_{222} = \frac{8}{15} \right\}, \left\{ \alpha_{221} = \frac{8}{15}, \alpha_{222} = \frac{7}{15} \right\} \right\} \quad (\text{III.55})$$

Pour la première solution nous trouvons $\beta(t_{11})=30$ et $\beta(t_{23})=7$, pour la deuxième $\beta(t_{11})=30$ et $\beta(t_{23})=8$. En procédant de la même manière, nous trouvons que la charge de la ressource la plus chargée dans les deux cas donnée par $19 \cdot 14 \text{ u.t.} = 266 \text{ u.t.}$ Puisque cette valeur est inférieure à $C^{15} = 270 \text{ u.t.}$, les ressources de type R3 ne sont pas critiques et le débit $\Phi_{\alpha} = \frac{1}{x} = 0.278 \text{ pièces/u.t.}$ peut être atteint.

◆ Alternativement, en utilisant les valeurs de α_{221} et α_{222} (III.53) obtenues pour un horizon $v_{\lambda\alpha}=10$, nous sommes dans les conditions d'application du résultat III.32 qui donne une condition suffisante pour la réalisation effective du débit. En effet, nous connaissons tous les éléments du routage α et l'horizon $v_{\lambda\alpha}$. Seul le type de ressource R3 doit être considéré avec $m_r=2$. Deux opérations, représentées par les transitions t_{11} et t_{23} , sont associées à ce type de ressource. Avec ces paramètres, la relation III.32 devient

$$\zeta = \text{ppcm} \left(\frac{m_3}{\text{pgcd}(m_3, \beta(t_{11}), \beta(t_{23}))} \right) = \frac{2}{\text{pgcd}(2, 20, 5)} = 2 \quad (\text{III.56})$$

Par conséquent un horizon $v = \zeta v_{\lambda\alpha} = 20$ est suffisant. Comparé à la valeur de l'horizon obtenu précédemment $v = 15$, ce résultat met en évidence le fait que la relation III.32 n'indique qu'une valeur suffisante pour l'horizon.

III.4.3.2 Minimisation de l'horizon

Si nous imposons une valeur v_0 pour l'horizon, il faut que tous les taux de routage α_{ijk} , $k=1..K_{ij}$ d'un conflit c_{ij} s'expriment avec des dénominateurs $q_{ij}=v_0 f_i$. Pour $v_0=1$ et $f_1 = 2$, $f_2 = 1$ et $f_3 = 2$, nous obtenons les valeurs admissibles suivantes

$$(\alpha_{111}, \alpha_{112}) \in \left\{0, \frac{1}{2}, 1\right\}^2 \quad (\alpha_{211}, \alpha_{212}) \in \{0, 1\}^2 \quad (\alpha_{221}, \alpha_{222}) \in \{0, 1\}^2 \quad (\text{III.57})$$

Il existe alors 12 combinaisons admissibles. Par un parcours exhaustif de l'espace de recherche nous trouvons que le plus grand débit effectivement réalisable est donné par

$$\alpha_{111} = \alpha_{112} = \frac{1}{2} \quad \alpha_{211} = 1 \quad \alpha_{212} = 0 \quad \alpha_{221} = 1 \quad \alpha_{222} = 0 \quad (\text{III.58})$$

Pour ce routage, le débit est imposé par le type de ressource R3. Les ressources de type R3 doivent effectuer 2 fois l'opération associée à t_{11} et une fois l'opération associée à t_{23} et nous avons donc $B_1(3) = \{t_{11}, t_{11}, t_{23}\}$. Ces trois opérations ont la même durée $\mu(t_{11}) = \mu(t_{23}) = 14$ u.t. La répartition de ces trois opérations de même durée sur deux ressources entraîne donc un fort déséquilibre entre les charges des deux ressources. Nous obtenons pour le CEF un temps de cycle $C^1 = 28$ u.t., ce qui correspond à un débit de $\Phi = 0.179$ pièces/u.t.

Dans le tableau suivant nous avons représenté les durées des cycles élémentaires virtuels (cf III.4) effectivement atteignables pour différentes valeurs de l'horizon.

v	1	2	3	4	5	6	7	8	9	10
C_v	28	21	18.7	19.5	19.6	18.7	18.6	18.3	18.7	18.2

v	11	12	13	14	15	16	17	18	19	20
C_v	18.4	18.3	18.2	18.4	18.0	18.3	18.2	18.1	18.3	18.0

Fig. III-17 Valeurs du couple débit-horizon

Ce tableau met en évidence le fait que la relation $C_v = f(v)$ n'est pas nécessairement *monotone*, ce qui complique l'analyse du problème. En particulier lorsque nous voulons traiter le cas de la maximisation du débit sous la contrainte $v \leq v_0$ (cf. III.2.3), nous ne pouvons affirmer que la solution se trouve sur une frontière du domaine considéré.

III.5 Conclusion

Les résultats présentés à la fin de ce chapitre sont basés sur l'analyse de trois types de problèmes.

Premièrement, nous avons identifié et défini l'horizon de réalisation comme troisième indice de performance d'un fonctionnement répétitif, complétant ainsi les deux autres indices généralement considérés, débit et en-cours. Nous avons alors justifié la prise en compte de cet indice dans le cadre de l'élaboration d'une commande déterministe d'un fonctionnement répétitif pour un système contenant d'une part des ressources de type identique et d'autre part des flexibilités au niveau des gammes.

Deuxièmement, pendant la construction du modèle RdP, nous avons mis en évidence la problématique qui consiste à caractériser certains types de flexibilités imbriquées ou enchaînées. Nous avons alors fait le choix de nous limiter à une modélisation pliée des flexibilités enchaînées et de n'admettre qu'un seul niveau d'imbrication dans la perspective de faciliter une optimisation. Les conséquences de ce choix sont cependant plus larges. En particulier, nous ne pouvons envisager une extension de l'approche aux classes D1 et D2. En effet, dans ce cas les taux de routage au niveau des flexibilités apparaîtraient alors sous forme de décisions dépendantes des ratios et nous serions de nouveau confronté à la problématique de choix imbriqués.

Troisièmement, le choix d'optimiser en premier lieu le couple débit-horizon sous l'hypothèse d'un en-cours suffisant a largement conditionné la façon d'appréhender le problème. Ce choix a en particulier permis de justifier l'hypothèse du découplage des ressources.

Nous avons alors pu mettre en équation le problème de l'optimisation conjointe du débit et de l'horizon et énoncer des conditions nécessaires et suffisantes pour la réalisation effective d'un fonctionnement répétitif. Les résultats obtenus montrent alors que la valeur de l'horizon nécessaire pour réaliser un certain débit, notamment le débit maximal, peut constituer un obstacle pour la réalisation effective au niveau discret. Notre approche qui consiste à considérer l'horizon en même temps que le débit dans la phase d'optimisation de flux se trouve alors confirmée.

IV. Modélisation et élaboration d'un ordonnancement pour un fonctionnement répétitif

IV.1 Introduction

Dans le chapitre précédent nous avons étudié l'optimisation du couple débit-horizon pour un fonctionnement répétitif. Comme résultats de cette étape nous obtenons les valeurs numériques de ces deux indices, associées à un ou éventuellement plusieurs routages. S'il existe plusieurs routages qui correspondent au même débit et au même horizon, il faut décider si un seul routage est à retenir ou si différents choix sont effectivement à considérer dans la suite. Bien que ces différents routages soient équivalents en ce qui concerne le débit et l'horizon, ils engendrent des charges différentes sur les ressources. La phase d'analyse que nous proposons d'aborder ici, part donc d'hypothèses différentes qui peuvent conditionner la minimisation de l'en-cours et donc aboutir à des résultats différents.

Le choix entre plusieurs routages admissibles n'est pas la seule situation où différentes possibilités apparaissent. Dans notre démarche qui doit aboutir à la construction d'un graphe d'événements, nous rencontrons d'autres situations où il est impossible de connaître la meilleure solution sans développer l'analyse jusqu'à son terme final, c'est-à-dire déterminer un ordonnancement total et l'en-cours associé. La première problématique que nous étudions concerne *l'extraction de chemins de gamme* à partir des taux de routage donnés. Dans ce contexte nous évoquons également la complexité engendrée par le problème de *partition des charges* qui se pose lorsqu'il existe plusieurs partitions admissibles pour la charge associée à un type de ressource disponible en plusieurs exemplaires.

Une autre problématique est engendrée par la manière de boucler les gammes qui sont soumises à la même contrainte sur l'en-cours. A cet égard, nous comparons deux approches que nous appelons *parallélisation* et *séquentialisation de gammes*. Nous établissons des bornes inférieures sur l'en-cours associées à ces deux approches et montrons qu'une séquentialisation des gammes doit toujours être envisagée. Cette séquentialisation des gammes introduit alors une nouvelle problématique relative à la *composition des gammes*. Un autre type d'indéterminisme est en rapport avec la

permutation d'opérations, type de flexibilité que nous avons écarté lors de l'analyse précédente en raison de l'insensibilité du débit et de l'horizon vis-à-vis de cette flexibilité. Rappelons que nous avons 'masqué' d'éventuelles flexibilités de permutation en choisissant arbitrairement un routage. Après la phase d'analyse de flux, ce type de flexibilité doit être réintroduit dans le modèle.

Mise à part la façon de boucler les gammes, nous ne pouvons pas, a priori, évaluer l'influence des différents choix sur le niveau de l'en-cours. Dans le cas général, pour effectuer une optimisation rigoureuse, nous devrions donc considérer tous les indéterminismes évoqués ci-dessus pendant la détermination simultanée des séquences de passage des produits sur les ressources (D5) et du niveau de l'en-cours et de sa répartition (D6a et D6b). Au vue de la complexité d'une prise en compte simultanée de tous les indéterminismes nous examinons les différents aspects d'une manière successive.

Nous présentons les différentes classes d'indéterminismes dans l'ordre évoqué ci-dessus, parce qu'il est le plus à même d'illustrer les implications sur les transformations successives d'un modèle RdP. A la fin de la première partie de ce chapitre nous obtenons alors un modèle ordonnançable, dans lequel il subsiste seulement des indéterminismes en rapport avec les classes D5 et D6. Pour cette dernière problématique nous présentons une heuristique qui permet de réaliser effectivement un fonctionnement répétitif spécifié par un couple débit-horizon. Pendant la construction de l'ordonnancement, l'algorithme tente de minimiser l'en-cours. Le résultat final est alors un graphe d'événements associé au fonctionnement répétitif.

IV.2 Construction d'un modèle ordonnançable

IV.2.1 Extraction des chemins de gamme et partition des charges

En présence de flexibilités d'affectation ou de procédé enchaînées, la modélisation pliée que nous avons retenue, ne permet pas toujours de déterminer les séquences d'opérations effectives par la seule connaissance des taux de routage (cf. III.3.1). Puisque les associations entre les choix retenus au niveau des différentes flexibilités sont

équivalentes en ce qui concerne les charge induites sur les ressources, nous ne pouvons exprimer aucune préférence pour telle ou telle association.

Rappelons que les taux de routage α_{ijk} sont indicés par $i=1..I$, $j=1..J_i$ et $k=1..K_{ij}$, désignant le $k^{\text{ème}}$ choix du $j^{\text{ème}}$ conflit de la $i^{\text{ème}}$ gamme. Si nous ne tenons pas compte d'éventuelles flexibilités de permutation, le nombre de chemins (suites déterministes d'opérations correspondant aux T-semiflots du RdP) pour fabriquer un produit de type i , noté b_i , est donné par

$$b_i = \prod_{j=1..J_i} K_{ij} \quad i=1..I \quad (\text{IV.1})$$

Puisque nous devons fabriquer vf_i produits de type i par cycle, le problème consiste à déterminer pour chaque type de produit $i=1..I$ un ensemble G_i contenant vf_i chemins. Ces chemins doivent être choisis de sorte que chaque transition t_j , représentative d'une opération, apparaisse exactement $\beta(t_j)$ fois dans cet ensemble, $\beta(t_j)$ étant obtenu par l'expression III.20. Nous désignons alors par $g(X_i^m)$, $i = 1..I$, $m = 1..b_i$, le nombre d'occurrences du chemin associé au T-semiflot X_i^m dans G_i . Les éléments d'un ensemble G_i doivent donc vérifier les deux relations suivantes

$$\sum_{m=1..b_i} g(X_i^m) = vf_i \quad i=1..I \quad (\text{IV.2})$$

$$\sum_{t \in |X_i^m|} g(X_i^m) = \beta(t) \quad \forall t \in T \text{ tq } \mu(t) > 0 \quad (\text{IV.3})$$

La relation IV.2 exprime le fait qu'il faut sélectionner exactement vf_i chemins pour chaque type de produit i tandis que IV.3 limite les choix à des combinaisons qui contiennent les transitions dans les quantités $\beta(t_j)$ imposées par le routage α retenu. Ces contraintes doivent alors être vérifiées pour toutes les gammes, avec ou sans flexibilité d'affectation ou de procédé enchaîné.

La seule différence dans ces deux cas réside dans le fait qu'en présence de flexibilités d'affectation ou de procédé enchaînés, il peut y avoir plusieurs ensembles G_i vérifiant les contraintes.

Considérons de nouveau l'exemple de la figure III-16 et plaçons nous dans les conditions d'un fonctionnement sur un horizon minimal, étudiées en III.4.3, avec les paramètres suivants:

$$\begin{aligned}
 f_1=2 & \quad f_2=1 & \quad f_3=2 & \quad v=1 & \quad C^1=28 \text{ u.t.} & \quad \Phi_\alpha=0.179 \text{ pièces/u.t.} \\
 \alpha_{111}=0.5 & \quad \alpha_{211}=1 & \quad \alpha_{221}=1 & \quad \beta(t_6)=1 & \quad \beta(t_{11})=2 & \quad \beta(t_{20})=0 & \quad \beta(t_{26})=0 & \quad \beta(t_{33})=2 \\
 \alpha_{112}=0.5 & \quad \alpha_{212}=0 & \quad \alpha_{222}=0 & \quad \beta(t_9)=1 & \quad \beta(t_{17})=1 & \quad \beta(t_{23})=1 & \quad \beta(t_{31})=2
 \end{aligned}$$

En appliquant IV.2 et IV.3 nous trouvons que les 5 produits de l'ensemble E_1 doivent être fabriqués selon les chemins de gamme suivants:

$$\begin{aligned}
 G_1 &= \{ X_1^1, X_1^2 \}, \text{ deux produits de type 1, l'un par } X_1^1 \text{ et l'autre par } X_1^2 \\
 G_2 &= \{ X_2^1 \}, \text{ un produit de type 2 par } X_2^1 \\
 G_3 &= \{ X_3^1, X_3^1 \}, \text{ deux produits de type 3 en utilisant le même chemin } X_3^1
 \end{aligned}$$

Dans notre exemple, malgré la présence de deux flexibilités d'affectation enchaînées, les valeurs numériques sont telles qu'il existe un seul ensemble G_2 satisfaisant les contraintes. Le RdP se transforme comme indiqué sur la figure IV-1. Après cette étape d'extraction de chemins de gamme, les flexibilités associées aux classes D3b et D3c peuvent alors être considérées comme définitivement résolues.

Pour obtenir le RdP de la figure IV-1, nous avons explicité les décisions en rapport avec la classe D4. En effet, pour fixer les valeurs du débit et de l'horizon nous avons seulement supposé qu'il existe une partition de la charge associée aux types de ressources. Il faut alors détailler cette partition. Dans le cas général, il est possible que nous soyons confronté à un grand nombre de possibilités qui sont a priori équivalentes.

Dans notre exemple, cette problématique se limite à la répartition de la charge associée aux deux ressources de type R3. Rappelons que pour les paramètres indiqués ci-dessus, cette charge est donnée par $B_1(3) = \{ t_{11}, t_{11}, t_{23} \}$ avec $\mu(t_{11}) = \mu(t_{23}) = 14$ u.t. (cf. III.4.3). Plusieurs partitions conviennent pour un temps de cycle de 28 u.t. Celle que nous avons retenue consiste à affecter les deux opérations associées à la transition t_{11} à la ressource R3.1 et l'opération associée à t_{23} à la ressource R3.2. Ce choix se justifie par le fait que chaque ressource est alors dédiée à un type de produits, ce qui peut simplifier la configuration des machines.

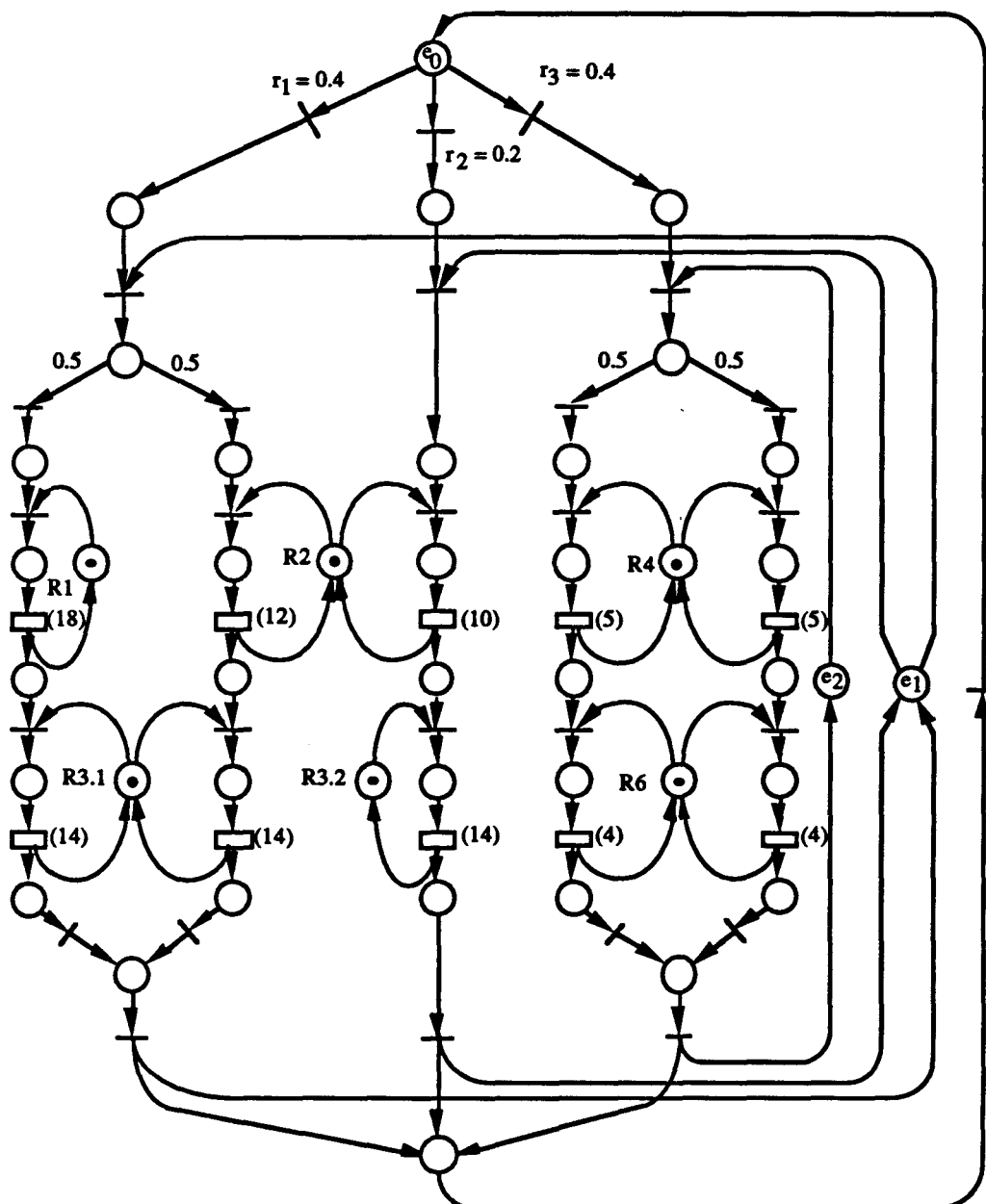


Fig. IV-1 Modèle après extraction des chemins et répartition des charges

Les opérations sont maintenant associées à des ressources individuelles. Le qualificatif 'restreint' associé aux gammes opératoires et introduit pour souligner qu'une opération a seulement été associée à un type de ressource, n'a plus de raison d'être. De même, les flexibilités d'affectation et de procédé existant dans les GORT ont été résolues ce qui peut être interprété comme une linéarisation partielle des gammes. Pour être plus

précis, nous pourrions caractériser ces nouvelles gammes par le terme '*gammes opératoires temporisées partiellement linéarisées*' (GOTPL). Cependant, pour simplifier la terminologie, nous parlerons dans la suite simplement de *gammes*.

Au niveau des places modélisant des conflits entre gammes d'un même type de produit, nous avons indiqué sur la figure IV-1 des taux de routage exprimant une résolution de conflit équitable (0.5 et 0.5). En effet, puisque nous avons dupliqué les gammes autant de fois que nécessaire pour satisfaire les contraintes IV.2 et IV.3, chaque gamme doit apparaître exactement une fois par cycle de fonctionnement.

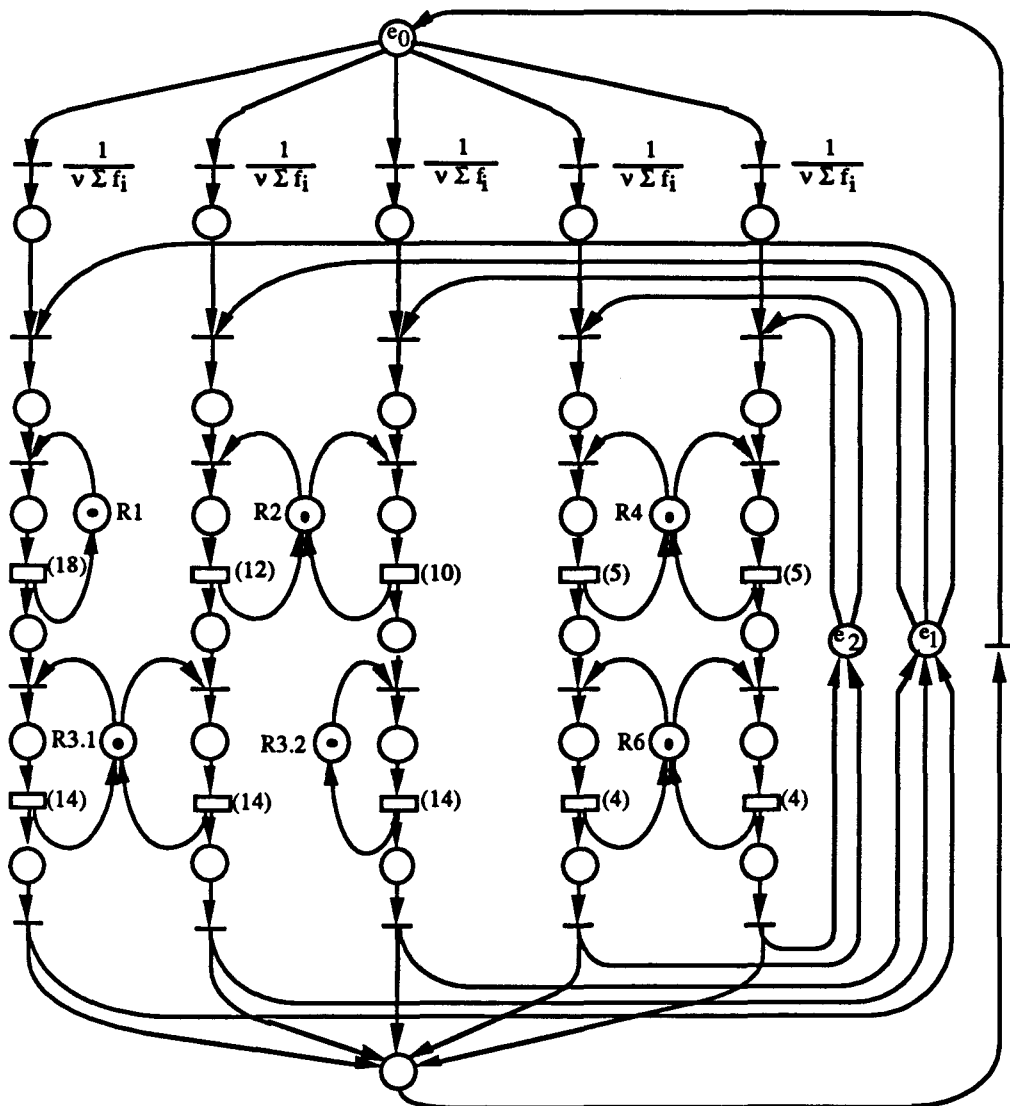


Fig. IV-2 RdP avec bouclage simplifié

De ce fait, l'existence de deux conflits au niveau du bouclage, un premier pour imposer les ratios de production et un deuxième pour une répartition équitable, n'est plus nécessaire. Nous pouvons donc simplifier le bouclage en utilisant un seul conflit avec des choix ayant des taux de routage identiques et égaux à $1/(v\Sigma f_i)$ comme indiqué sur la figure IV-2.

Remarquons que la répartition des marques e_0 , e_1 et e_2 peut être a priori quelconque. Le marquage de la figure IV-2 n'est pas nécessairement un marquage atteignable par un fonctionnement répétitif de débit Φ_α et d'horizon v . En effet, nous ne pouvons préciser les valeurs de e_0 , e_1 et e_2 et la répartition des marques qu'après avoir établi l'ordonnement. Dans la section suivante nous allons d'abord étudier une autre transformation de la structure du modèle, qui a une influence sur le niveau de l'en-cours.

IV.2.2 Parallélisation et séquentialisation des gammes

Dans le RdP obtenu par les transformations évoquées ci-dessus, trois types de conflits subsistent. Le premier concerne l'ordre dans lequel les ressources exécutent les opérations qui leur sont affectées. Il s'agit donc de la problématique d'ordonnement proprement dit. Le deuxième type de conflit se situe au niveau des places de bouclage intérieur. Il est relatif à la façon de distribuer l'en-cours parmi les gammes concernées. Le troisième type de conflit concerne le bouclage extérieur modélisant le respect des ratios entre les gammes.

A ces trois types d'indéterminismes s'ajoutent d'éventuelles flexibilités de permutation de l'ordre des opérations d'une gamme (D3a), non représentées dans notre modèle. Ces flexibilités de permutation peuvent faciliter la tâche d'ordonnement si elles peuvent être incluses dans l'algorithme d'ordonnement. L'heuristique que nous allons présenter à la fin de ce chapitre ne permet pas d'en tenir compte et nous devons par conséquent procéder comme nous l'avons fait en présence de routages équivalents, de partitions de charges équivalentes et d'ensembles de chemins de gamme équivalents: nous retenons un choix qui ne peut être évalué qu'en menant l'analyse jusqu'à son terme. Si le résultat (le niveau de l'en-cours) n'est pas satisfaisant, alors nous revenons sur le choix effectué et explorons une autre branche de l'arborescence.

Afin de pouvoir évaluer la qualité d'un ordonnancement établi, il est intéressant de disposer de bornes inférieures du niveau de l'en-cours. Un résultat, c'est-à-dire un ordonnancement avec l'en-cours associé, peut alors être comparé à cette borne. La décision d'itérer ou non, en explorant une autre branche de l'arborescence, se trouve ainsi facilitée. Dans la suite nous allons établir des bornes inférieures de l'en-cours selon deux approches différentes de résolution des conflits au niveau des places de bouclage interne, comme présenté dans [OHL 95]. Pour effectuer cette comparaison, nous avons besoin d'introduire quelques définitions.

Définition 11: La durée associée à un T-semiflot X , notée $\mu(X)$, est la somme des durées des transitions appartenant au support de X .

$$\mu(X) = \sum_{t \in |X|} \mu(t) \quad (\text{IV.4})$$

Définition 12: Une *macro-gamme* est une concaténation de gammes (de type GOTPL). Une macro-gamme est dite *homogène*, si elle est constituée de gammes identiques. Sinon elle est dite *hétérogène*.

La durée associée à une macro-gamme est alors donnée par la somme des durées des gammes qui la composent.

IV.2.2.1 Parallélisation des gammes

Une première approche pour résoudre les conflits au niveau des places de bouclage intérieur consiste à boucler chaque gamme individuellement. Cette approche consiste à partitionner l'en-cours total en $v \sum f_i$ sous-ensembles qui sont chacun dédiés à une gamme. Nous pouvons alors évaluer l'en-cours minimal nécessaire pour un fonctionnement cyclique.

En effet, nous savons que chaque gamme bouclée constituera un circuit de fabrication* du modèle graphe d'événements que nous voulons construire. Nous notons γ_X ce circuit contenant toutes les transitions temporisées de la gamme associée au T-semiflot X .

* nous utilisons ici la terminologie introduite en II.4.5

Le temps de cycle associé à γ_X , noté C_{γ_X} est alors donné par le rapport entre la somme des durées des transitions et le nombre de marques dans γ_X (cf. II.4.4).

$$C_{\gamma_X} = \frac{\mu(X)}{M(\gamma_X)} \quad (\text{IV.5})$$

Pour valider notre hypothèse selon laquelle l'en-cours ne constitue jamais un facteur limitatif pour un fonctionnement répétitif avec un temps de cycle de C^v , il faut donc que $C_{\gamma_X} \leq C^v$. Si nous désignons par $\lceil \cdot \rceil$ l'opérateur qui donne le plus petit entier supérieur ou égal à l'argument, l'en-cours dans chaque gamme bouclée doit être supérieur ou égal à une certaine valeur donnée par

$$M_{\min}(\gamma_X) = \left\lceil \frac{\mu(X)}{C^v} \right\rceil \quad (\text{IV.6})$$

Par conséquent, en faisant la sommation sur toutes les $v \sum_{i=1..I} f_i$ gammes, l'en-cours total est borné inférieurement par

$$B_0 = \sum_{i=1..I} \left(\sum_{X \in G_i} \left\lceil \frac{\mu(X)}{C^v} \right\rceil \right) \quad (\text{IV.7})$$

IV.2.2.2 Séquentialisation des gammes

Au lieu de boucler les gammes individuellement, nous pouvons regrouper des gammes associées à la même contrainte sur l'en-cours S_h , $h = 1..H$. Ces regroupements peuvent s'effectuer de différentes manières. En effet, il faut décider d'une part *quelles* sont les gammes à associer et d'autre part *comment* ces différentes gammes sont associées. Sur le dessin de la figure IV-3 nous avons représenté de façon schématique les H ensembles contenant respectivement les gammes associées à la contrainte S_h , $h = 1..H$. Ces ensembles, désignés par G^h , sont donnés par

$$G^h = \bigcup_{i \in S_h} G_i \quad h = 1..H \quad (\text{IV.8})$$

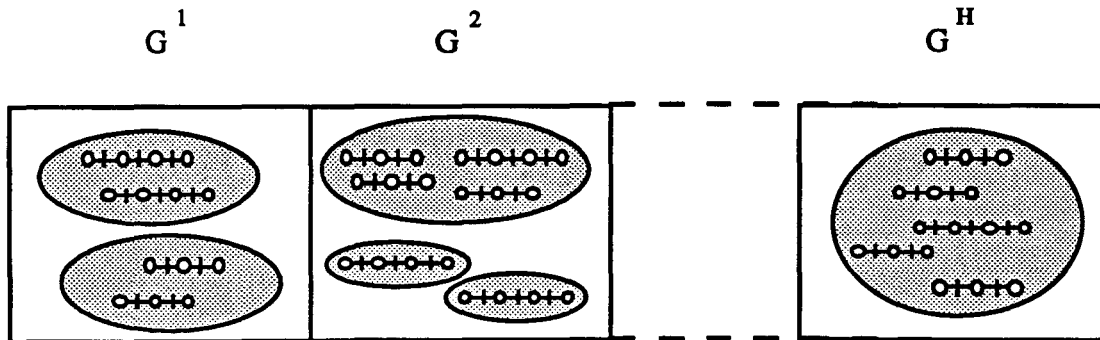


Fig. IV-3 Partition de l'ensemble des gammes

Les éléments d'un ensemble G^h , $h = 1..H$, peuvent alors être associés pour former des macro-gammes. Nous appelons \mathcal{P}_h une partition d'un ensemble G^h comme représenté sur la figure IV-3.

La complexité de cette approche se situe à deux niveaux. Premièrement, le nombre de partitions différentes croît très vite avec la cardinalité des ensembles G^h et avec le nombre de types de gammes différents que ces ensembles contiennent. Deuxièmement, pour une partition donnée, les macro-gammes peuvent être conçues de différentes manières.

En effet, si un sous-ensemble est composé de plus de trois types de gammes différents, ou de 2 types différents qui apparaissent chacun au moins deux fois, nous pouvons obtenir plusieurs macro-gammes différentes.

Les figures IV-4, IV-5 et IV-6 illustrent cette situation.

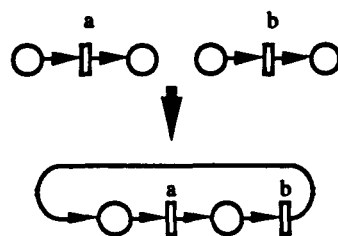


Fig. IV-4 Association de deux gammes différentes

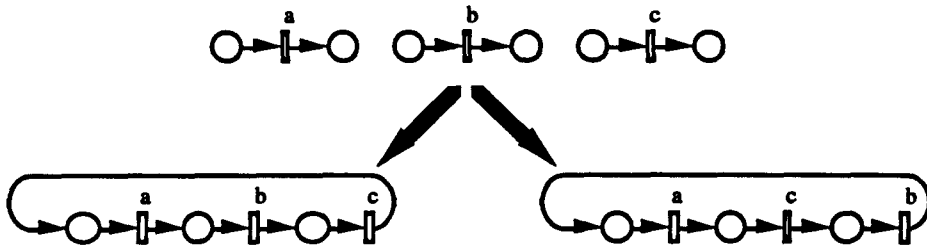


Fig. IV-5 Associations de trois gammes différentes

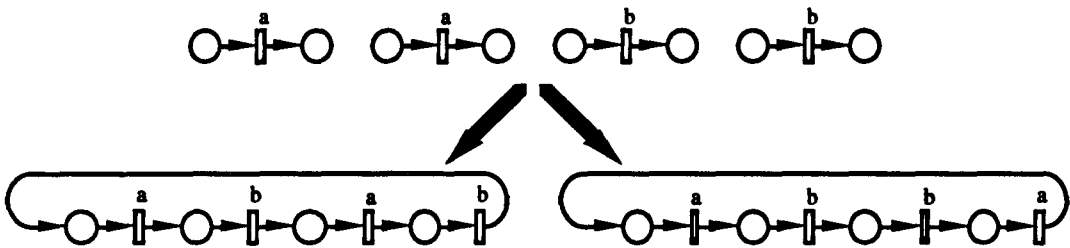


Fig. IV-6 Associations des quatre gammes appartenant à deux types différents

Evidemment, avec un nombre de gammes croissant, la combinatoire engendrée augmente très vite. Si nous désignons alors par M_a^h , $a = 1..A_h$, la $a^{\text{ème}}$ macro-gamme issue de la partition \mathcal{P}_h de G^h , nous pouvons associer une nouvelle borne à l'en-cours, qui s'exprime pour un type de contrainte $h = 1..H$ par

$$B_h(\mathcal{P}_h) = \sum_{a=1..A_h} \left[\frac{\mu(M_a^h)}{C^v} \right] \quad h = 1..H \quad (\text{IV.9})$$

Si nous désignons par \mathcal{P} l'union des partitions \mathcal{P}_h , la sommation sur l'indice h donne la borne pour l'en-cours total

$$B(\mathcal{P}) = \sum_{h=1..H} \left(\sum_{a=1..A_h} \left[\frac{\mu(M_a^h)}{C^v} \right] \right) \quad (\text{IV.10})$$

Dans cette expression, comme dans celle de B_0 , (cf. IV.7), les sommations portent sur l'ensemble des transitions temporisées du modèle RdP de la figure IV-2. La différence entre les bornes pour l'en-cours total provient seulement de la manière de comptabiliser les durées associées aux transitions. A cause de l'inégalité

$$\lceil a \rceil + \lceil b \rceil \geq \lceil a + b \rceil \quad \forall (a, b) \in IQ^+ \times IQ^+ \quad (IV.11)$$

nous avons donc pour toute partition \mathcal{P}

$$B_0 \geq B(\mathcal{P}) \quad (IV.12)$$

En tenant compte de la relation IV.11, il est clair que la meilleure borne est obtenue pour une partition \mathcal{P}^* qui consiste à regrouper toutes les gammes d'un ensemble G^h dans une seule macro-gamme. Dans ce cas $A_h = 1$ pour tout $h = 1..H$. La durée associée à la macro-gamme engendrée est alors donnée par

$$\mu(M_1^h) = \sum_{i \in S_h} \left(\sum_{X \in G_i} \mu(X) \right) \quad h = 1..H \quad (IV.13)$$

Par conséquent, l'expression IV.10 se transforme en

$$B^*(\mathcal{P}^*) = \sum_{h=1..H} \left[\sum_{i \in S_h} \left(\sum_{X \in G_i} \frac{\mu(X)}{C^v} \right) \right] \quad (IV.14)$$

Cette dernière borne donne alors la plus petite valeur que nous puissions espérer pour le niveau de l'en-cours. Nous avons par conséquent

$$B_0 \geq B(\mathcal{P}) \geq B^*(\mathcal{P}^*) \quad (IV.15)$$

Remarquons ici que les ensembles G^h , $h = 1..H$, et par conséquent \mathcal{P}_h , \mathcal{P} et $B(\mathcal{P})$ sont sensibles à la façon d'extraire les chemins de gammes. Seule la borne B^* est insensible à l'extraction, car elle est obtenue par l'arrondi de la somme des durées de toutes les opérations soumises à la même contrainte. Par conséquent B^* peut être déterminée par la seule connaissance du routage α^2 retenu.

IV.2.2.3 Atteignabilité des bornes

Nous abordons à présent la question de savoir d'une part, si les inégalités IV.15 peuvent être strictes et d'autre part, si les bornes établies peuvent effectivement être atteintes. Afin de répondre à cette question et de comparer les différentes approches de bouclage des gammes, nous considérons un exemple simple qui permet d'illustrer les résultats obtenus.

Nous supposons que trois types de produits, A, B et C doivent être fabriqués par une seule ressource avec des ratios 0.25, 0.25 et 0.5. Chaque gamme contient une seule opération et toutes les durées opératoires sont prises égales à l'unité. L'ensemble minimal de pièces est $E_1 = \{A, B, C, C\}$ et nous tentons de réaliser un fonctionnement répétitif avec $C^1 = 4$ u.t. Nous supposons de plus que les produits de type A et B sont associés à un même type d'en-cours, tandis que les produits de type C requièrent un autre type d'en-cours. Nous partons donc des gammes comme indiqué sur la figure IV-7.

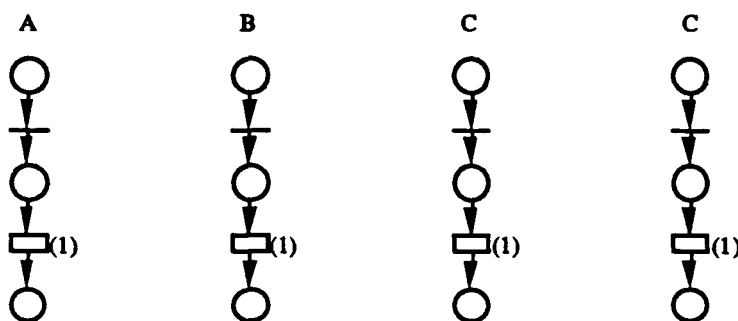


Fig. IV-7 Gammes pour des ratios de production 0.25, 0.25 et 0.5

Avec l'approche de parallélisation des gammes, en utilisant l'expression IV.7, nous obtenons la valeur de la borne B_0

$$B_0 = \left\lceil \frac{1}{4} \right\rceil + \left\lceil \frac{1}{4} \right\rceil + \left\lceil \frac{1}{4} \right\rceil + \left\lceil \frac{1}{4} \right\rceil = 4 \quad (\text{IV.16})$$

Un en-cours de 4 est effectivement suffisant, comme l'indique le graphe d'événements de la figure IV-8, représentatif d'un fonctionnement répétitif avec un temps de cycle $C^1 = 4$ u.t.

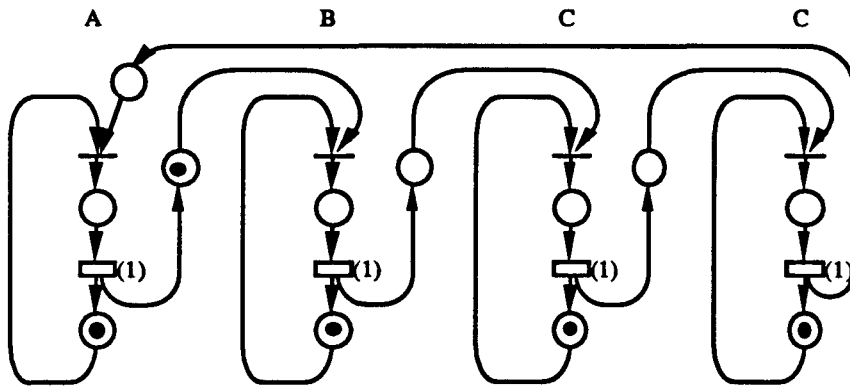


Fig IV-8 Parallélisation des gammes

Considérons maintenant une partition \mathcal{P} qui consiste à regrouper les deux gammes de type C dans une macro-gamme et qui garde les gammes associées aux types de produits A et B isolées. En utilisant l'expression IV.10, nous obtenons alors

$$B(\mathcal{P}) = \left\lceil \frac{1}{4} \right\rceil + \left\lceil \frac{1}{4} \right\rceil + \left\lceil \frac{2}{4} \right\rceil = 3 \quad (\text{IV.17})$$

Cette borne est atteignable, car un en-cours de 3 est suffisant comme l'indique le graphe d'événements de la figure IV-9, représentatif d'un fonctionnement répétitif avec $C^1 = 4$ u.t.

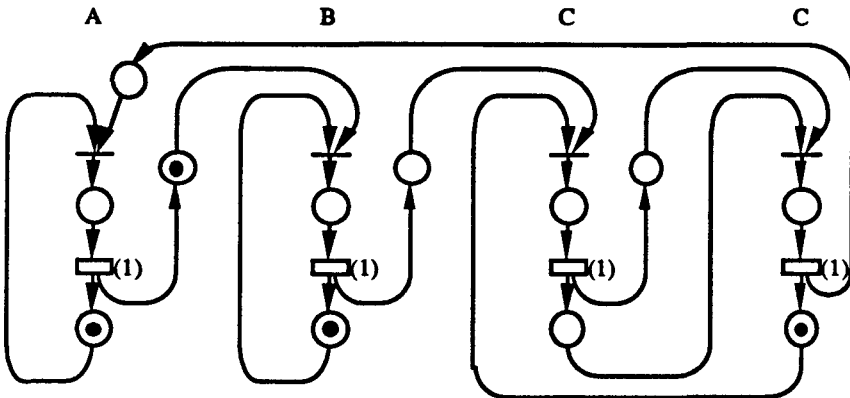


Fig. IV-9 Séquentialisation par type de produit*

* les graphes d'événements des figures IV-9 et IV-10 contiennent des places implicites qui ne sont pas nécessaires pour le fonctionnement. Elles jouent cependant un rôle explicatif, en ce qui concerne les contraintes que le bouclage des gammes peut induire sur le niveau de l'en-cours, c'est la raison pour laquelle nous ne les supprimons pas.

Si nous considérons maintenant la partition \mathcal{P}^* , nous devons regrouper dans une macro-gamme toutes les gammes associées à une même contrainte. Par conséquent, les gammes des produits A et B sont associées. En utilisant l'expression IV.14, nous obtenons alors

$$B^*(\mathcal{P}^*) = \left\lceil \frac{2}{4} \right\rceil + \left\lceil \frac{2}{4} \right\rceil = 2 \quad (\text{IV.18})$$

Cette borne est également atteignable, comme l'indique le graphe d'événements de la figure IV-10, représentatif d'un fonctionnement répétitif avec $C^1 = 4$ u.t.

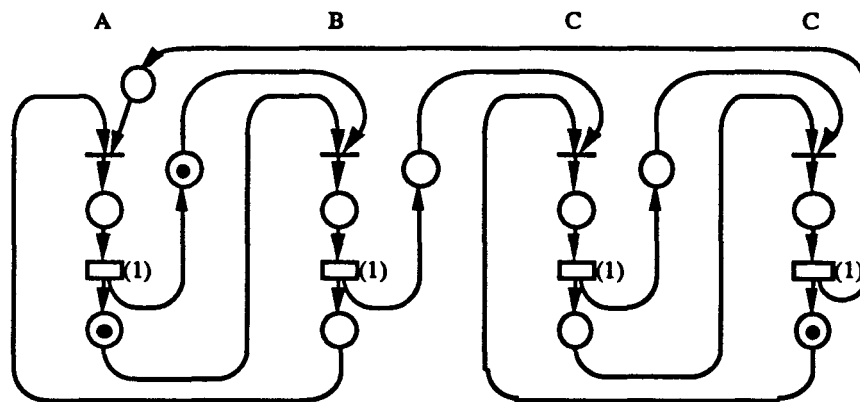


Fig IV-10 Séquentialisation par type de contrainte

Les bornes établies précédemment sont donc atteignables pour l'exemple que nous venons de considérer. Pour savoir si les bornes sont atteignables dans tous les cas de figure, nous devons considérer les différentes classes de circuits élémentaires d'un graphe d'événements modélisant un fonctionnement répétitif. En effet, les circuits élémentaires ont été partagés en trois classes distinctes (cf II.4.5):

- circuits de commande, modélisant l'ordonnancement des ressources
- circuits de fabrication, correspondant aux gammes bouclées
- circuits mixtes, passant par des places des deux autres classes de circuits

Puisque nous voulons que le plus lent des circuits de commande impose le temps de cycle, nous devons faire en sorte que les deux autres types de circuits ne ralentissent

pas le fonctionnement. Les considérations précédentes sur l'en-cours nécessaire pour différents bouclages des gammes ne concernaient de fait que les circuits de fabrication.

Tant que l'ordonnancement n'a pas été fixé nous ne connaissons pas les circuits mixtes. Par conséquent, à ce stade de l'analyse, il est impossible de décider de l'atteignabilité des bornes dans le cas général. Dans certains cas nous pouvons cependant affirmer que les bornes sont atteignables. En effet, l'existence d'exemples sans circuits mixtes prouve l'atteignabilité des bornes. (cf. exemple des figures IV-9 et IV-10 mais aussi figure III-1).

IV.2.2.4 Comparaison des deux approches

L'approche par séquentialisation des gammes que nous venons d'étudier apparaît alors comme une généralisation de l'approche par parallélisation. En effet, le bouclage individuel des gammes n'est qu'un cas particulier de l'approche par séquentialisation qui correspond à la partition engendrant un nombre maximal de gammes bouclées.

Nous avons vu qu'à cause de l'existence des circuits mixtes, nous ne pouvons déterminer a priori la partition qui donnera effectivement le meilleur en-cours. Cependant, nous avons montré que différentes partitions engendrent des bornes différentes pour l'en-cours et que ces bornes sont atteignables dans certains cas.

Les conséquences de ces remarques sont importantes:

◆ Premièrement, la séquentialisation des gammes doit impérativement être prise en compte pendant la transformation du modèle. Si nous nous limitons à un bouclage individuel des gammes, nous explorons seulement une petite partie de l'espace de recherche et nous écartons en particulier le sous-espace qui contient la ou les partitions associées à la borne B^* . La limitation à l'approche par parallélisation peut alors engendrer une dégradation de la valeur de l'en-cours, indice de performance que nous cherchons à optimiser. En effet, cette dégradation risque d'être d'autant plus marquée que le nombre de gammes élémentaires est grand. Rappelons que le nombre de gammes correspond à la cardinalité de E_v , qui est proportionnelle à v et aussi à $\sum_{i=1..I} f_i$ (cf. III.2.1).

◆ Deuxièmement, en ce qui concerne la manière d'aborder le problème, il découle de l'analyse précédente que l'espace de recherche doit être parcouru selon des partitions engendrant des bornes B croissantes, c'est-à-dire depuis les partitions correspondant à la plus petite borne B^* jusqu'aux partitions associées au plus grand B_0 . En effet, si pendant le parcours, le meilleur en-cours trouvé dans la partie du parcours déjà effectuée est inférieur ou égal aux bornes associées aux branches qui restent à parcourir, nous pouvons arrêter le parcours car nous sommes sûr que dans la partie restante de l'arborescence, les résultats obtenus ne sont pas meilleurs. En revanche, il faut parcourir tout l'espace de recherche, si la meilleure solution est supérieure ou égale au plus grand B_0 .

IV.2.2.5 Application à l'exemple

Considérons de nouveau l'exemple de la figure IV-2. Les gammes sont soumises à deux types de contraintes. En appliquant les résultats établis précédemment, nous pouvons déterminer les bornes. En tenant compte des valeurs indiquées sur la figure IV-2, nous obtenons

$$B_0 = \left\lceil \frac{18+14}{28} \right\rceil + \left\lceil \frac{12+14}{28} \right\rceil + \left\lceil \frac{10+14}{28} \right\rceil + \left\lceil \frac{5+4}{28} \right\rceil + \left\lceil \frac{5+4}{28} \right\rceil = 6 \quad (\text{IV.19})$$

$$B^* = \left\lceil \frac{18+14+12+14+10+14}{28} \right\rceil + \left\lceil \frac{5+4+5+4}{28} \right\rceil = 4 \quad (\text{IV.20})$$

Selon la méthode indiquée à la fin de la section IV.2.2.4, nous commençons la recherche avec les partitions associées à la borne $B = 4$. Il existe ici *deux* partitions:

La première est la partition \mathcal{P}^* ayant servi pour obtenir l'expression IV.20, regroupant les trois gammes associées à la contrainte S_1 et les deux gammes associées à la contrainte S_2 . La deuxième partition consiste à regrouper les gammes en trois sous-ensembles contenant respectivement, la première et la troisième gamme associées à la contrainte S_1 , la deuxième gamme associée à la contrainte S_1 et les deux gammes associées à la contrainte S_2 . En effet, avec cette dernière partition, nous obtenons aussi

$$B^* = \left\lceil \frac{18+14+10+14}{28} \right\rceil + \left\lceil \frac{12+14}{28} \right\rceil + \left\lceil \frac{5+4+5+4}{28} \right\rceil = 4 \quad (\text{IV.21})$$

Puisque la première de ces deux partitions contient un regroupement de trois gammes différentes, elle engendre *deux* macro-gammes différentes (cf. figure IV-5). En somme, nous devons donc considérer trois possibilités, qui sont a priori équivalentes, car associées à la même borne $B = B^* = 4$. Nous procédons alors comme nous l'avons fait précédemment : nous retenons un choix parmi les trois et considérons les autres possibilités seulement si ce choix s'avère insatisfaisant dans la suite. Nous avons représenté le choix que nous retenons sur la figure IV-11.

Sur cette figure apparaissent également les ressources associées aux transformations ainsi que les transitions immédiates entre les gammes pour souligner l'existence d'un chargement-déchargement de pièces. Avec une autre représentation nous pourrions perdre la trace des gammes initiales composant les macro-gammes. En revanche, le bouclage commun relatif au respect des ratios n'apparaît plus dans le modèle.

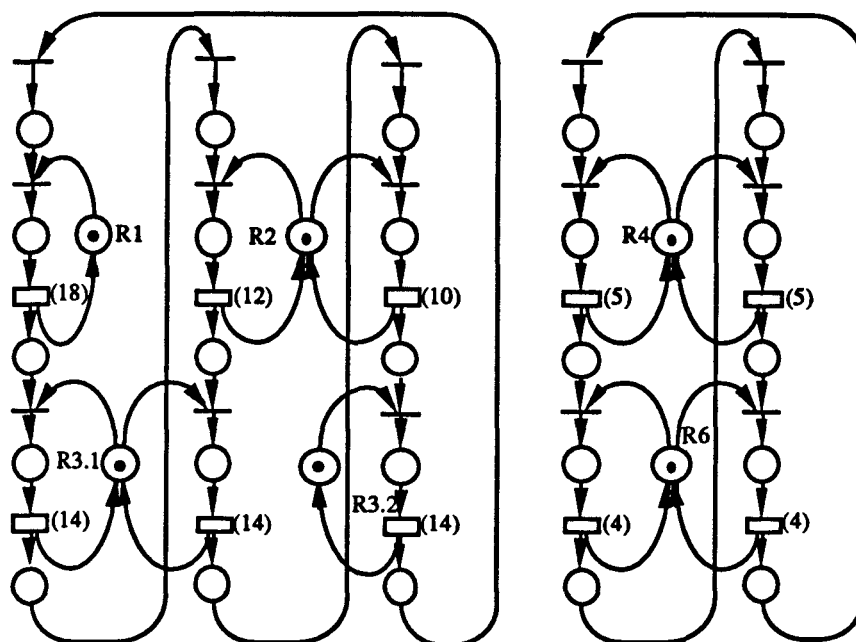


Fig. IV-11 Macro-gammes hétérogène et homogène

IV.2.3 Contraintes pour un fonctionnement cyclique

Le fait d'avoir enlevé le bouclage extérieur n'altère pas l'analyse, car nous n'avons plus besoin d'imposer les ratios de production. En effet, nous avons extrait autant de gammes que nécessaire pour réaliser le fonctionnement répétitif et nous savons

alors que chaque gamme doit être utilisée une et une seule fois par cycle de fonctionnement. D'autre part nous avons vu que l'en-cours dans les gammes peut être considéré comme un moyen pour relaxer les contraintes de précedence stricte des gammes. Cette interprétation est relative à un cycle de fonctionnement car un produit particulier doit bien évidemment suivre l'ordre des opérations imposé par les gammes.

Pour un fonctionnement répétitif nous devons alors imposer un tir *unique* de chaque transition par cycle. En ce qui concerne les RdP, l'analyse de cette situation fait appel à la notion de *distance synchronique* entre des tirs de transitions [SIL 87][SIL 92], induisant l'idée d'une dépendance stricte entre les fréquences de tirs d'un sous-ensemble de transitions. De plus, l'heuristique d'ordonnement que nous allons présenter dans la suite utilise une première simplification qui introduit une *restriction* pour l'ensemble des ordonnancements réalisable : elle suppose un fonctionnement par cycles disjoints, c'est-à-dire *l'absence de chevauchement d'opérations* entre cycles adjacents, tel que nous l'avons spécifié au début du chapitre II. Pour la transformation de notre modèle RdP, nous allons tenir compte de cette contrainte supplémentaire due à l'heuristique d'ordonnement.

IV.2.3.1 Modélisation de la contrainte

Avec un dispositif comme celui de la figure IV-12, il est possible d'imposer un respect des taux de routage pour un nombre limité de décisions en gardant toutes les possibilités de permutation.

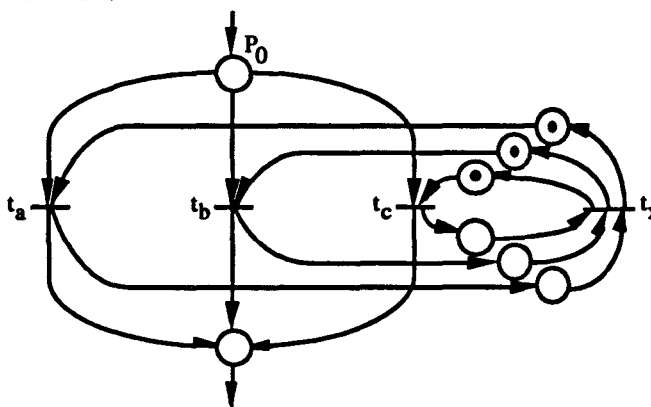


Fig IV-12 Modélisation d'un respect de taux de visite

Des dispositifs de ce type ont été étudiés dans un contexte plus général dans [OHL 94a]. Avec des poids associés aux arcs et avec un nombre de marques variable, ils

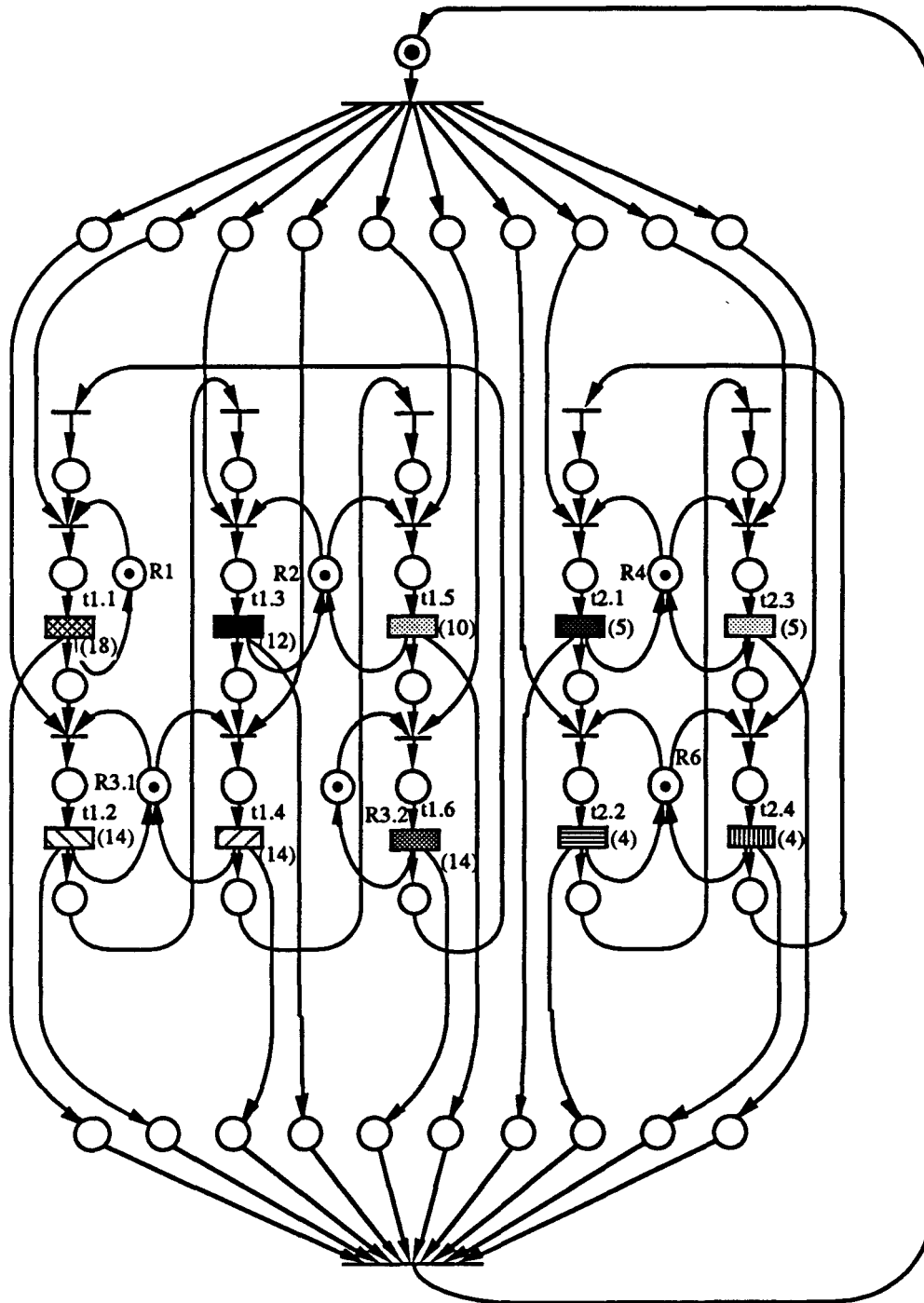


Fig. IV-13 Modèle ordonnançable

permettent d'imposer des taux de routage quelconques avec des possibilités de permutation variables.

Afin d'imposer un et un seul tir de chaque transition par cycle de fonctionnement, ainsi que des cycles disjoints, nous introduisons alors un tel dispositif dans notre modèle RdP. Pour simplifier les notations et faciliter la présentation de l'algorithme dans la section suivante, nous avons renuméroté les transitions représentatives des opérations et associé une couleur à chaque transition comme indiqué sur la figure IV-13. Nous qualifions le modèle obtenu après cette transformation de *modèle ordonnançable*. En effet, les seuls conflits restant sont ceux relatifs aux classes D5 et D6.

IV.3 Ordonnancement et en-cours

IV.3.1 Position du problème

Nous disposons donc d'un modèle ordonnançable, la dernière phase de la transformation en graphe d'événements peut alors être appréhendée de différentes manières. Une première possibilité consiste à fixer arbitrairement un ordonnancement. Sur la structure du graphe d'événements ainsi obtenu, il s'agit alors de déterminer l'en-cours nécessaire pour que le temps de cycle souhaité soit effectivement respecté. Cette approche a été développée dans [HIL 89a][LAF 91][LAF 92] où des algorithmes de minimisation de marques dans un graphe d'événements ont été présentés. Dans les termes de notre classification, ceci revient à séparer la résolution des conflits associés aux classes D5 et D6.

Dans [MUN 91] un ordonnancement cyclique a été obtenu qui respecte le temps de cycle imposé par les ressources critiques. D'éventuelles contraintes de précedence incompatibles avec des contraintes d'accès aux ressources sont relaxées en retardant les opérations d'un certain nombre de cycles de fonctionnements. Cette technique correspond en fait au placement d'en-cours dans le système, où encore de marques dans le modèle RdP. Puisque ces travaux se situent dans un contexte informatique, l'en-cours et l'optimisation de l'en-cours ne sont pas pris en compte explicitement. Remarquons que

dans ce contexte particulier, l'en-cours correspond aux emplacements mémoire pour stocker les résultats intermédiaires.

La résolution conjointe des deux classes de décision D5 et D6 est en fait difficile. Dans [HIL 87b][HIL 89a] une heuristique a été présentée qui place l'en-cours pendant la phase d'ordonnancement. Malheureusement, cette heuristique n'arrive pas toujours à respecter le temps de cycle imposé. En effet, pendant sa mise en œuvre, il est possible que l'algorithme bloque à un certain niveau. Dans ce cas il faut augmenter le temps de cycle et itérer. Dans certains cas le temps de cycle du graphe d'événements final est alors allongé de 15% ou plus. Il est toutefois possible d'utiliser la seule structure du graphe d'événements obtenu et de ré-appliquer sur cette structure un des algorithmes de minimisation de marques cités ci-dessus pour obtenir finalement le temps de cycle imposé.

Afin d'éviter une approche avec un nombre d'itérations a priori inconnu ainsi qu'un ré-ajustement de l'en-cours après la phase d'ordonnancement, nous proposons dans la suite un algorithme qui respecte un temps de cycle donné [OHL 94b]. Dans la suite nous notons C^* ce temps de cycle à respecter.

IV.3.2 Algorithme d'ordonnancement

L'algorithme que nous avons développé est une heuristique basée sur le principe d'une construction progressive. Sa principale originalité réside dans le fait qu'il s'appuie pleinement sur la dualité entre ressources et gammes. Pour expliquer le principe de l'algorithme, nous introduisons les définitions suivantes:

Définition 13: La *marge de ressource initiale* est la différence entre C^* et la charge associée à cette ressource.

Définition 14: La *marge de gamme initiale* est la différence entre C^* et la somme des durées des opérations de cette gamme.

Après la phase d'initialisation, l'ordonnancement est construit de manière progressive en plaçant une opération par itération. Pour placer une opération, un arbre d'une profondeur π est parcouru, arbre qui contient toutes les séquences possibles de π

prochaines opérations. Un coût est associé à chaque séquence prévisionnelle et la première opération de la séquence de coût minimal est ordonnancée. Pendant la construction d'une séquence prévisionnelle, toutes les opérations sont placées au plus tôt. La date de début d'une opération est alors donnée par la plus grande des dates de disponibilité du produit et de la ressource associés à cette opération. Supposons que ces deux dates ne coïncident pas et appelons Δ la différence. Deux cas peuvent se produire:

(1) Si Δ est dû à l'indisponibilité de la ressource, le produit reste inutilisé pendant Δ en attendant la ressource. L'opération peut seulement commencer à la date de disponibilité de la ressource et nous perdons Δ de la marge de gamme, sans perdre de la marge de ressource.

(2) Si Δ est dû à l'indisponibilité du produit, la ressource reste inutilisée pendant Δ en attendant le produit. Si nous plaçons l'opération à la date de disponibilité du produit, nous perdons Δ de la marge de ressource sans perdre de la marge de gamme.

Puisque nous ne considérons pas le cas où des opérations peuvent être insérées ultérieurement dans des intervalles d'attente des ressources, toute consommation de la marge d'une ressource est irrévocable. Il apparaît alors que si nous voulons respecter C^* , il faut absolument éviter qu'une marge de ressource ne devienne négative.

Si dans le cas (2), la marge de ressource est inférieure à Δ , nous ne pouvons placer l'opération à la date de disponibilité du produit, parce que dans ce cas la marge de ressource deviendrait négative. En introduisant un autre produit, disponible immédiatement, nous pouvons éviter cette perte de marge de ressource. Nous pouvons toujours procéder de cette manière, ce qui nous permet de garantir que C^* sera respecté. Puisque nous voulons minimiser l'en-cours, il faut cependant veiller à ajouter un produit seulement lorsque c'est effectivement incontournable.

L'algorithme place alors l'opération avec éventuellement un nouvel en-cours. Les marges sont mises à jour et l'algorithme itère.

Au vue de la description de l'algorithme il est possible de dégager les paramètres de la fonction de coût pour une évaluation des différentes séquences prévisionnelles. Une fonction de coût doit pénaliser l'ajout d'en-cours et aussi la consommation des marges de

ressource et des marges de gamme. Puisqu'un grand nombre d'opérations à placer nécessite généralement une marge plus importante, nous avons pondéré dans le calcul du coût d'une séquence, la marge de chaque ressource et de chaque gamme par le nombre d'opérations restantes.

IV.3.3 Description de l'algorithme

IV.3.3.1 Paramètres et variables

Paramètres de l'algorithme:

J	nombre de gammes, $J = \sum_{h=1..H} A_h$ avec $\forall \sum_{i=1..I} f_i \geq J \geq H$
M	nombre de ressources, $M \leq \sum_{r=1..R} m_r$
t_{jk}	$k^{\text{ème}}$ opération de la $j^{\text{ème}}$ gamme, $j = 1..J$, $k = 1..K_j$
$\mu(t_{jk})$	durée associée à t_{jk}
$m(t_{jk})$	ressource associée à t_{jk} , $m(t_{jk}) \in \{1..M\}$, $j = 1..J$, $k = 1..K_j$
C^*	temps de cycle à respecter
π^*	profondeur maximale de l'arbre de recherche
$c_{\text{en-cours}}$	coût associé au placement d'un en-cours
c_{gamme}	coût associé à la perte d'une unité de marge de gamme
$c_{\text{ressource}}$	coût associé à la perte d'une unité de marge de ressource

Variables de l'algorithme:

π	profondeur effective de l'arbre de recherche
$n(j)$	en-cours dans la gamme, $j = 1..J$
$O_G^-(j)$	nombre d'opérations restantes pour la gamme, $j = 1..J$
$O_M^-(m)$	nombre d'opérations restantes pour la ressource, $m = 1..M$
$O_G^+(j)$	nombre d'opérations de la gamme j déjà placées, $j = 1..J$ vérifiant alors $O_G^-(j) + O_G^+(j) = K_j$

$\delta_G(j)$	date de disponibilité de l'en-cours dans la gamme, $j = 1..J$
$\delta_M(m)$	date de disponibilité de la ressource, $m = 1..M$
$\epsilon_G(j)$	marge de la gamme, $j = 1..J$
$\epsilon_M(m)$	marge de la ressource, $m = 1..M$
$\tau_G(j)$	temps opératoire restant pour la gamme, $j = 1..J$
	donné par $\tau_G(j) = \sum_{k=(O_{G^+}(j)+1)..K_j} \mu(t_{jk})$
$\tau_M(m)$	temps opératoire restant pour la ressource, $m = 1..M$
	donné par $\tau_M(m) = \sum_{\substack{m(t_{jk})=m \\ k > O_{G^+}(j)}} \mu(t_{jk})$

La figure IV-14 illustre la relation entre les différents paramètres et variables*. Elle représente la situation juste après l'étape d'initialisation qui vient de placer respectivement un en-cours devant les opérations associées à t1.1 et t2.1.

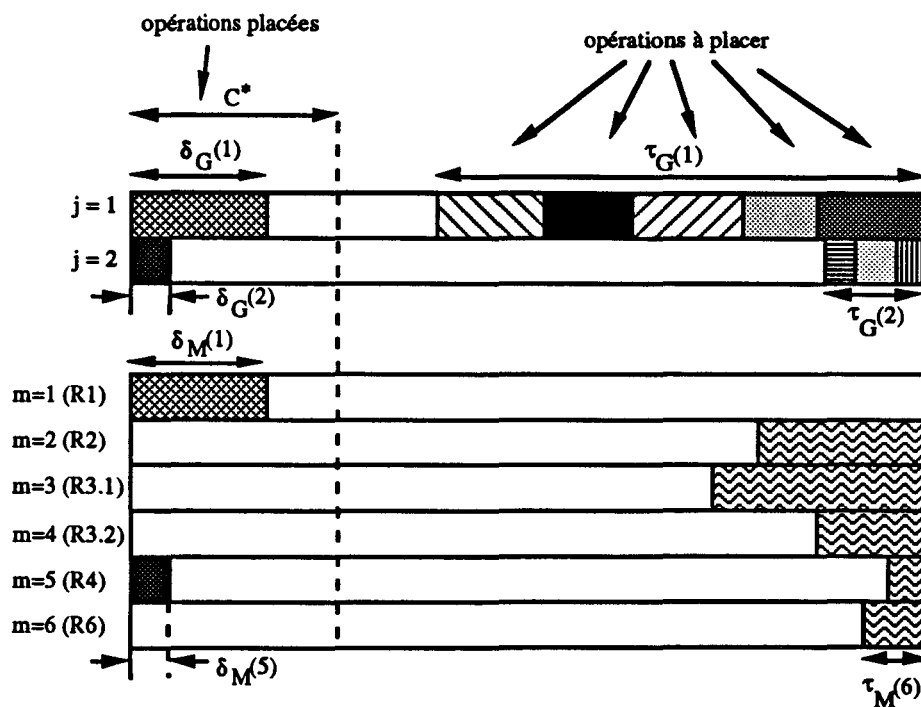


Fig. IV-14 Paramètres de l'algorithme

* en ce qui concerne les opérations restantes au niveau des ressources, l'ordre de ces opérations est inconnu et nous pouvons seulement représenter le temps total $\tau_M(m)$.

Nous définissons alors les marges de la manière suivante:

$$\varepsilon_G(j) = C^* - \tau_G(j) - \delta_G(j) \quad j = 1..J \quad (\text{IV.22})$$

$$\varepsilon_M(m) = C^* - \tau_M(m) - \delta_M(m) \quad m = 1..M \quad (\text{IV.23})$$

Nous pouvons remarquer que les marges des gammes peuvent être négatives dès le départ. Ceci signifie que nous devons obligatoirement ajouter des en-cours lors du déroulement de l'algorithme.

En ce qui concerne l'initialisation, il est difficile d'exprimer un meilleur choix a priori et nous supposons donc dans un premier temps une initialisation quelconque qui consiste à placer une marque dans chaque macro-gamme, sachant que nous pouvons revenir sur ce choix s'il s'avère insatisfaisant dans la suite.

Afin de limiter la profondeur effective de l'arbre de recherche par le nombre total d'opérations à placer, nous définissons la profondeur effective de l'arbre de recherche par

$$\pi = \min \left(\pi^*, \sum_{j=1..J} O_{G^-}(j) \right) \quad (\text{IV.24})$$

IV.3.3.2 Séquences de placement admissibles

Soit σ une séquence de π opérations, nous désignons par $\sigma(i)$ la $i^{\text{ème}}$ opération de la séquence et σ^m la restriction de σ aux opérations effectuées par la ressource m . Nous avons alors

$$\sigma = t_{j_1 k_1} \circ t_{j_2 k_2} \circ t_{j_3 k_3} \circ \dots \circ t_{j_\pi k_\pi} \quad (\text{IV.25})$$

$$\sigma(i) = t_{j_i k_i} \quad i = 1.. \pi \quad (\text{IV.26})$$

$$\sigma^m = \text{projection}(\sigma / m(t_{j_k}) = m) \quad m = 1..M \quad (\text{IV.27})$$

Nous appelons $\omega(j, \sigma)$ l'ensemble des indices des opérations de σ qui appartiennent à la gamme j . Nous désignons respectivement par $\min(\omega(j, \sigma))$, $\max(\omega(j, \sigma))$ et $\text{card}(\omega(j, \sigma))$ le plus petit élément, le plus grand élément et le cardinal de cet ensemble.

L'ensemble des séquences de longueur π admissibles, noté $\Omega(\pi)$, est alors donné par

$$\Omega(\pi) = \left\{ \sigma / \forall j = 1..J, \forall (a, b) \in \omega^2(j, \sigma), (a < b) \Rightarrow (k_a < k_b), \right. \\ \left. (k_{\min(\omega(j, \sigma))} = O_{G^+}(j) + 1), (k_{\max(\omega(j, \sigma))} = O_{G^+}(j) + \text{card}(\omega(j, \sigma))) \right\} \quad (\text{IV.28})$$

Les trois conditions que chaque couple d'opérations réperé par les indices a et b dans une séquence prévisionnelle doit vérifier sont les suivantes:

(i) Si deux opérations de la même gamme j apparaissent dans σ respectivement aux positions a et b , telles que a est avant b , il faut retrouver ce même ordre au niveau de la gamme j .

(ii) Il faut que la première opération dans σ relative à la gamme j soit effectivement la prochaine opération à placer pour la gamme j .

(iii) S'il existe plusieurs opérations dans σ relatives à la gamme j , il faut que celles-ci représentent une suite ininterrompue d'opérations de la gamme j .

Deux séquences $(\sigma_1, \sigma_2) \in \Omega^2(\pi)$ sont alors équivalentes, si elles vérifient

$$\sigma_1^m = \sigma_2^m \quad m = 1..M \quad (\text{IV.29})$$

En effet, la seule différence entre deux séquences qui vérifient IV.29 peut se trouver au niveau de l'ordre dans lequel les opérations effectuées par des ressources différentes apparaissent. Les résultats des placements des opérations des deux séquences sont alors identiques. Par conséquent, il suffit de considérer un ensemble $\Omega'(\pi)$ contenant un et un seul représentant de chaque classe d'équivalence.

IV.3.3.3 Procédure de placement d'une opération

L'algorithme de placement fait appel à la procédure de placement que nous décrivons ici. Cette procédure place au plus tôt la $k^{\text{ème}}$ opération de la $j^{\text{ème}}$ gamme. Si la marge de ressource $\delta_M(t_{jk})$ est suffisante, l'opération est placée directement à la date au plus tôt. Sinon, un nouvel en-cours est ajouté et l'opération commence à la date de disponibilité de la ressource. Ensuite les différents paramètres sont mis à jour.

placer(t_{jk})

- 1) si $(\delta_G(j) - \delta_M(m(t_{jk}))) > \varepsilon_M(m(t_{jk}))$ alors {incrémenter $n(j)$;
faire $\delta_G(j) := 0$ }
- 2) faire $x := \max\{\delta_G(j), \delta_M(m(t_{jk}))\}$
- 3) faire $\delta_G(j) := x + \mu(t_{jk})$
- 4) faire $\delta_M(j) := x + \mu(t_{jk})$
- 5) décrémenter $O_{G-(j)}$
- 6) décrémenter $O_{M-(m(t_{jk}))}$
- 7) incrémenter $O_{G+(j)}$
- 8) faire $\tau_G(j) := \tau_G(j) - \mu(t_{jk})$
- 9) faire $\tau_M(m(t_{jk})) := \tau_M(m(t_{jk})) - \mu(t_{jk})$
- 10) faire $\varepsilon_G(j) = C^* - \tau_G(j) - \delta_G(j)$
- 11) faire $\varepsilon_M(m(t_{jk})) = C^* - \tau_M(m(t_{jk})) - \delta_M(m(t_{jk}))$

La mise à jour concerne alors respectivement la disponibilité de la gamme (3) et de la ressource (4), le nombre d'opérations restantes pour la gamme (5) et pour la ressource (6), le nombre d'opérations effectuées de la gamme (7), le temps opératoire restant pour la gamme (8) et pour la ressource (9) et, finalement, la marge de gamme (10) et la marge de ressource (11).

IV.3.3.4 Algorithme de placement

Nous avons alors réuni tous les éléments pour donner l'algorithme d'ordonnement:

tant qu'il existe une opération à placer, ($\sum_{j=1..J} O_{G^-}(j) > 0$), faire

1) déterminer la profondeur π par IV.24

2) calculer l'ensemble des séquences à tester $\Omega'(\pi)$

3) choisir la séquence σ^* de plus petit coût:

$\forall \sigma \in \Omega'(\pi)$ faire

3.1) pour $i := 1.. \pi$ placer($\sigma(i)$)

3.2) Coût_1 := $c_{\text{en-cours}} \times \sum_{j=1..J} n(j)$

3.3) Coût_2 := $c_{\text{ressource}} \times \sum_{m=1..M} O_{M^-}(m) \times \epsilon_M(m)$

3.4) Coût_3 := $c_{\text{gamme}} \times \sum_{j=1..J} O_{G^-}(j) \times \epsilon_G(j)$

3.5) Coût_total := Coût_1 - Coût_2 - Coût_3

4) placer($\sigma^*(1)$)

Avant l'exécution de l'algorithme il faut fixer les valeurs de la profondeur maximale π^* ainsi que les paramètres de la fonction de coût. Puisque les pertes de marges de ressources sont irrécupérables, nous leurs donnons une plus grande importance. De même, il faut fortement pénaliser l'ajout d'un-cours. De bons résultats ont été obtenus avec les valeurs suivantes [OHL 94b]:

$$c_{\text{en-cours}} = 200 \quad c_{\text{ressource}} = 10 \quad c_{\text{gamme}} = 1$$

Ces valeurs sont données à titre indicatif et peuvent être adaptées à des cas particuliers. Ainsi il est possible de définir des paramètres propres aux différentes gammes. En ce qui concerne la profondeur maximale π^* , nous pouvons utiliser ce paramètre pour ajuster le temps de calcul selon les contraintes imposées.

IV.3.4 Mise en œuvre

En appliquant l'algorithme à notre exemple, nous obtenons l'ordonnancement et la distribution de l'en-cours représentés sur la figure IV-15.

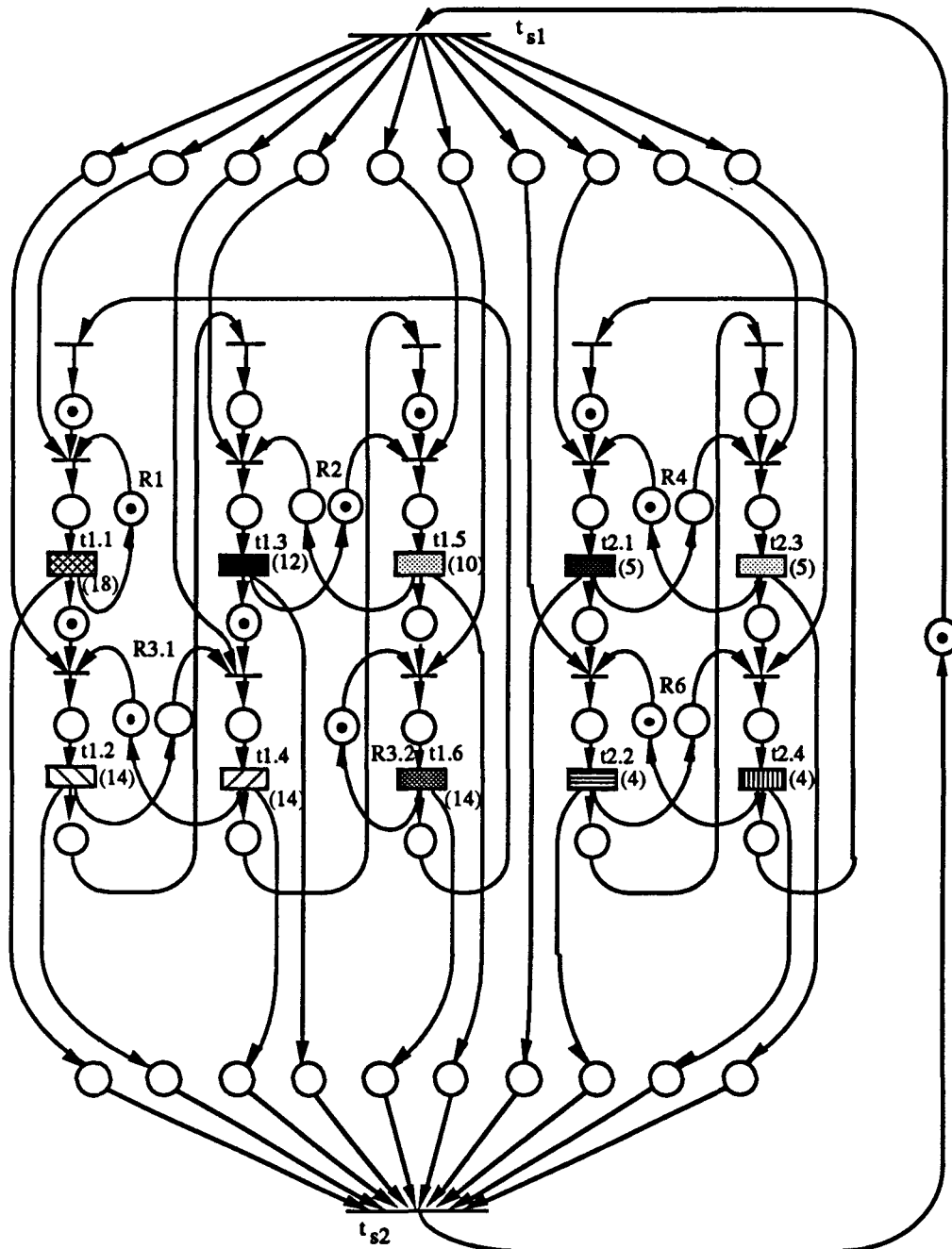


Fig. IV-15 Modèle ordonnancé contraint au fonctionnement par cycles disjoints

Nous obtenons alors un en-cours total de $n = 5$, valeur qui est supérieure à la borne $B^* = 4$ calculée précédemment, mais inférieure à la borne $B_0 = 6$. Pour interpréter ce résultat, il faut rappeler que d'un côté l'algorithme d'ordonnement utilisé est seulement une heuristique et d'un autre côté les bornes pour l'en-cours ne sont pas nécessairement atteignables. En effet, la différence entre la borne et le résultat obtenu peut être due à la sous-optimalité de l'algorithme ou à la non-atteignabilité de la borne.

Dans le cas général, si le niveau d'en-cours obtenu n'est pas satisfaisant, nous devons revenir sur les différents choix effectués auparavant et explorer les autres branches au niveau :

- ◆ des *flexibilités de permutation* des opérations
- ◆ des *compositions des macro-gammes* à partir des gammes élémentaires dans les sous-ensembles définis par les partitions retenues
- ◆ des *partitions de l'ensemble des gammes* soumises à une même contrainte sur l'en-cours
- ◆ de la *répartition des charges* associées aux types de ressources sur les ressources individuelles
- ◆ de l'*extraction de chemins de gammes* pour satisfaire le routage α retenu
- ◆ du *choix du routage α^2* parmi ceux qui correspondent au couple débit-horizon spécifié

Rappelons que nous avons cependant pu établir une condition d'arrêt pour le parcours de l'arborescence, si le parcours est ordonné selon des bornes croissantes (cf. IV.2.2.4). En ce qui concerne l'exemple illustratif, toutes les branches associées à des bornes plus grandes que $B = 4$ peuvent être éliminées après le premier résultat avec $n = 5$.

En l'absence d'un algorithme d'ordonnement *exact* s'appliquant aux macro-gammes bouclées non-contraintes à un fonctionnement par cycles disjoints (cf. figure IV-11) même un parcours complet de l'arborescence ne peut pas garantir une optimalité globale. C'est le cas de l'exemple traité pour lequel un ordonnancement optimal existe avec $n = 4$ mais ce dernier requiert des chevauchements d'opérations entre cycles adjacents et ne peut donc pas être obtenu avec l'heuristique proposée.

IV.4 Simplification du graphe d'événements

Le modèle obtenu après la transformation en graphe d'événements peut encore être simplifié. En effet, nous connaissons maintenant au moins une transition dont la date de fin de tir coïncide avec la date de fin de cycle. Nous choisissons alors une de ces transitions et utilisons l'événement caractérisant la fin de tir pour déclencher le cycle de fonctionnement suivant. Ceci nous permet de supprimer toutes les autres liaisons entre les transitions modélisant des opérations et la transition t_{s2} en bas du modèle de la figure IV-15.

Parmi les liaisons entre la transition t_{s1} en haut du modèle et les transitions modélisant le début des opérations, nous devons seulement conserver celles qui servent effectivement à retarder le début des opérations jusqu'à la date d'exécution prévue. Sur la figure IV-16 nous avons représenté l'ordonnancement d'un point de vue ressource et d'un point de vue en-cours. Pour cela nous identifions les marques dans le RdP et nous les désignons par J1 à J5 comme indiqué sur la figure IV-16.

Puisque nous avons opté pour un tir au plus tôt à l'intérieur d'un cycle de fonctionnement, seules les opérations qui commencent à la date de début de cycle sont susceptibles d'être avancées. Dans ce cas, une opération peut alors être avancée si la ressource et le produit sont tous les deux disponibles avant la date prévue.

Dans notre exemple, ces conditions sont remplies par les premières opérations associées aux ressources R1 et R4. Par conséquent, nous ne devons contrôler que le début de tir de ces deux transitions. Les autres liaisons peuvent ainsi être supprimées, sans que le fonctionnement ne se trouve modifié. Le graphe d'événements obtenu n'est pas nécessairement fortement connexe. Dans certains cas, nous pourrions même obtenir un graphe d'événements composé de parties disjointes.

Remarquons que le bouclage servant à séparer les cycles de fonctionnements peut engendrer une nouvelle classe de circuit s'ajoutant aux trois classes définies auparavant. Néanmoins, par construction, ces *circuits de retardement* respectent le temps de cycle imposé par la ou les ressource(s) critique(s).

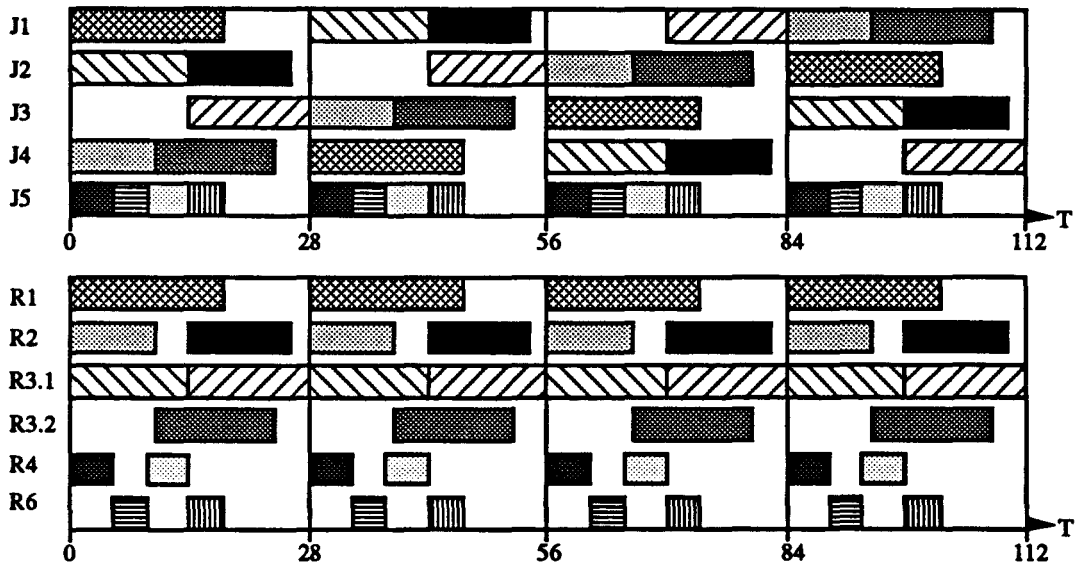
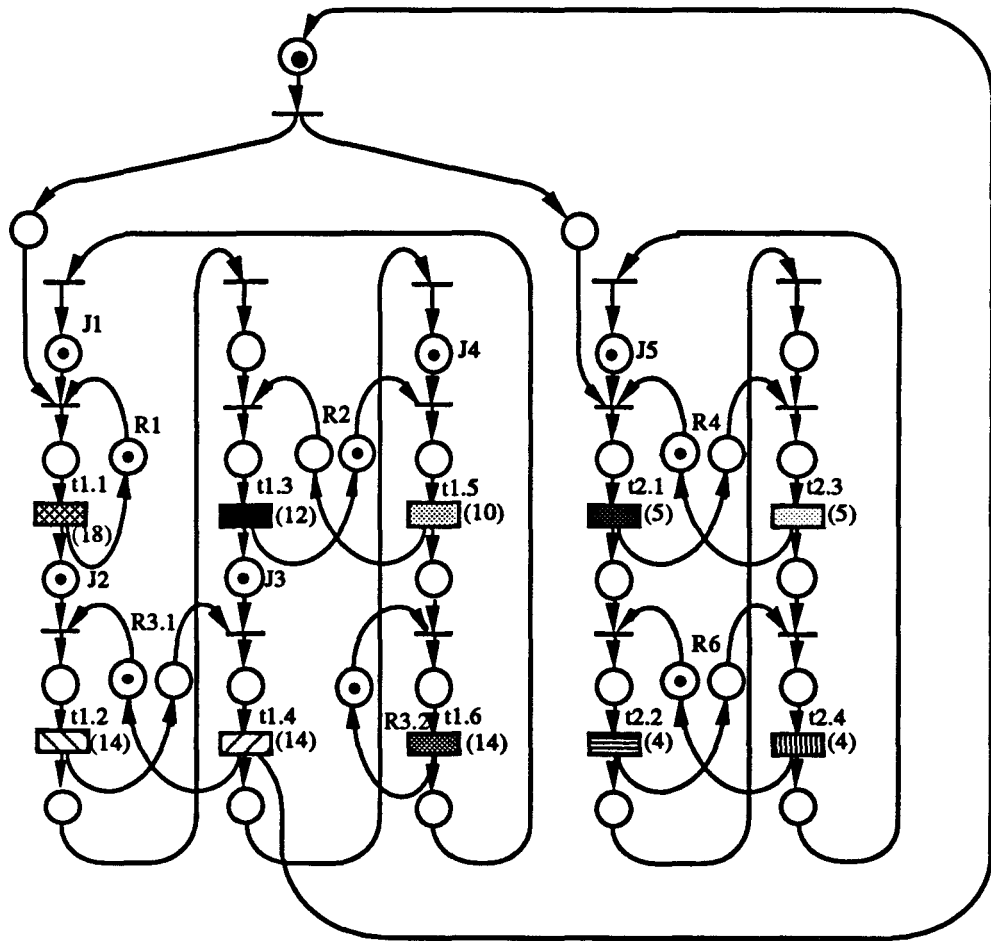


Fig. IV-16 Ordonnancement répétitif: vue de l'en-cours et des ressources

IV.5 Conclusion

Au cours de ce chapitre nous avons étudié les types de décisions en aval de l'analyse de flux. Nous avons en particulier défini l'étape d'extraction de chemins de gammes qui complète celle de la détermination des taux de routage pour la résolution des flexibilités d'affectation (D3b) et de procédé (D3c). De même, il a fallu détailler la répartition des charges associées aux types de ressources en plusieurs exemplaires. La résolution de ces classes de problèmes se décompose donc en deux étapes relatives respectivement à l'analyse au niveau flux et au niveau discret.

Les résultats que nous avons obtenus pour l'approche par séquentialisation des gammes entraînent deux conséquences. Premièrement, nous avons montré que la parallélisation des gammes peut induire des contraintes supplémentaires sur l'en-cours nécessaire pour atteindre un certain débit. Remarquons que l'approche par séquentialisation est intéressante, même si des flexibilités et des ressources multiples ne sont pas prises en compte. En effet, cette approche aussi est applicable dans le cas où l'on dispose de gammes linéaires et où ces gammes sont dupliquées pour respecter des ratios de production. Deuxièmement, l'analyse des bornes associées aux différentes partitions nous indique la meilleure façon de procéder pour appréhender la combinatoire engendrée par l'approche par séquentialisation. Par conséquent, nous pouvons parcourir l'espace de recherche selon un itinéraire qui permet de limiter éventuellement une énumération exhaustive.

L'algorithme d'ordonnancement que nous avons présenté à la fin de ce chapitre est une heuristique simple qui possède néanmoins un certain nombre de caractéristiques intéressantes. En premier lieu, l'algorithme aboutit toujours à un ordonnancement avec le temps de cycle imposé et respecte alors les contraintes imposées par la première phase d'optimisation du couple débit-horizon. De plus, le temps de calcul peut être adapté par le choix de la profondeur de l'arbre de recherche. Pour des exemples de petite taille, une recherche quasiment exhaustive peut alors être envisagée.

Finalement, les transformations successives effectuées sur le modèle RdP ont mis en évidence l'adéquation de cet outil d'accompagner un processus d'analyse tel que nous l'avons présenté.

V. Conclusion générale

V.1 Analyse des résultats

Les apports du travail que nous avons présenté dans ce mémoire concernent pour l'essentiel l'évaluation et l'optimisation de performances des systèmes flexibles de production manufacturière. Les résultats obtenus s'étendent cependant au contexte plus général de systèmes à événements discrets sous réserve des hypothèses de validité. Nous rappelons ci-après les principaux résultats présentés dans ce mémoire.

Tout d'abord, lors de l'extension de l'étude de fonctionnements répétitifs aux cas des ressources multiples et des flexibilités de différents types, nous avons mis en évidence la nécessité de tenir compte d'un nouvel indice de performance que nous avons appelé horizon de réalisation. En deuxième lieu, sous l'hypothèse d'un en-cours suffisant, nous avons étudié l'optimisation conjointe du débit de production et de l'horizon associé pour une réalisation au niveau discret. A cet effet, nous avons mis en équation les liens entre les taux de routage et ces deux indices de performance. Ceci nous a permis d'énoncer des conditions nécessaires et suffisantes pour une réalisation effective du débit et de l'horizon par un fonctionnement répétitif.

L'analyse de ces résultats, illustrée par l'étude d'un exemple, a permis de conclure à l'atteignabilité du débit de production maximal, à condition toutefois d'accepter un horizon éventuellement très long. En pratique, la valeur numérique de l'horizon associé au débit maximal peut cependant interdire la réalisation effective. Il apparaît donc qu'il faut essayer de maîtriser d'emblée ces deux indices simultanément dans une démarche d'analyse quantitative.

Après la phase d'optimisation de flux, nous avons détaillé les décisions restantes afin d'obtenir une commande déterministe avec un en-cours optimisé. Puisque la combinatoire engendrée par les différents types de décisions en interdit de fait la résolution simultanée, nous avons examiné ces décisions successivement. Pour les types de décisions qu'aucun critère ne permet de distinguer, nous avons proposé d'effectuer un choix a priori, sous réserve de revenir sur celui-ci ultérieurement. En revanche, en ce qui

concerne la problématique de l'association des gammes pour former des macro-gammes, nous avons mis en évidence des approches différentes et établi des bornes inférieures pour l'en-cours associé. Nous avons démontré que ces bornes sont atteignables dans certains cas et nous en avons déduit une méthode pour parcourir l'espace de recherche. Finalement, nous avons proposé une heuristique d'ordonnement qui transforme le modèle en un graphe d'événements représentatif d'un fonctionnement répétitif.

La décomposition du problème en différents types de décisions a mis en évidence la combinatoire potentielle. Son optimisation rigoureuse nécessiterait un parcours exhaustif de l'espace de recherche ainsi qu'un algorithme d'ordonnement pour lequel l'optimalité devrait être garantie. En pratique, nous avons proposé d'explorer seulement quelques combinaisons au niveau des différents choix, permettant néanmoins d'obtenir un résultat satisfaisant et proche de l'optimum.

Par ailleurs, il convient de se poser la question sur le bien-fondé d'une optimisation rigoureuse de l'en-cours. En effet, lors de la modélisation du système de transport et lors de la présentation de l'heuristique d'ordonnement nous avons mis en relief le rôle tampon de l'en-cours pour découpler les différentes ressources du système. Un système fonctionnant avec le strict minimum d'en-cours peut devenir trop sensible à des perturbations, mêmes de faibles intensités, dues par exemple à des durées de transport variables.

Cette remarque ne diminue en rien l'intérêt de l'approche par séquentialisation des gammes que nous proposons. En effet, l'approche par parallélisation de gammes introduit des contraintes sur l'en-cours par la nature même de la modélisation du problème. Dans ce contexte, l'en-cours supplémentaire se trouve alors confiné dans certaines parties du système en raison du bouclage des gammes. Rien ne permet d'affirmer a priori que cette localisation de l'en-cours soit la meilleure pour contribuer à atténuer réellement les effets de petites perturbations. Un éventuel accroissement de l'en-cours dans l'optique d'une plus grande robustesse du système doit se faire en augmentant des durées de transport prévisionnelles, ce qui entraîne alors implicitement non seulement une augmentation de l'en-cours mais aussi une répartition adaptée de celui-ci.

En fait, l'étude de la problématique que nous venons d'évoquer dépend de l'interprétation que nous attachons au graphe d'événements, résultat de notre démarche.

Si nous considérons qu'il s'agit d'un modèle de prévision du comportement temporel du système, nous devons tenir compte d'éventuelles petites variations. Ceci peut être le cas si d'autres événements (approvisionnement, évacuation de pièces finies) dépendent des dates prévues. En revanche, il est également possible d'interpréter notre modèle comme un modèle de commande de référence qui servira finalement à synchroniser l'évolution asynchrone du processus physique. Dans le cas d'éventuelles perturbations, le comportement temporel de notre système ne correspond alors plus exactement à celui attendu. Remarquons que dans cette perspective il serait également possible d'envisager de laisser fonctionner le système dans un mode au plus tôt afin de dégager la plus grande marge.

Les résultats que nous venons de présenter sont valables dans un contexte général. Bien que les hypothèses parfois restrictives utilisées dans ce mémoire limitent le champ d'application des résultats obtenus, ces résultats s'inscrivent dans une démarche cohérente d'analyse quantitative. Dans ce sens, ils s'inspirent de l'esprit de la démarche de conception CASPAIM. Les résultats d'ensemble de ce mémoire développent la composante d'analyse quantitative du projet et apportent ainsi une phase complémentaire selon une approche originale et relativement générale.

En particulier dans la phase de modélisation, nous avons pu nous appuyer sur des éléments de modélisation établis dans le projet. Les résultats obtenus dans ce mémoire corroborent un certain nombre de choix effectués dans le projet CASPAIM, notamment en ce qui concerne l'utilisation des Réseaux de Petri comme principal outil de modélisation ainsi que l'approche hiérarchique proposée antérieurement.

V.2 Perspectives de recherche

Plusieurs perspectives de recherche découlent de notre travail.

En premier lieu il paraît nécessaire de rechercher des algorithmes adaptés à la résolution de certains sous-problèmes identifiés dans ce mémoire. Cette remarque concerne notamment la détermination des taux de routage ainsi que celle de la répartition des charges associées à tout type de ressources multiples. En effet, les résultats obtenus pour ces sous-problèmes influencent directement les valeurs du couple débit-horizon. A

cet effet, une évaluation de la complexité de ces problèmes peut être utile afin de déterminer dans quelle mesure une éventuelle recherche exhaustive est envisageable.

Evoquons pour terminer quelques axes de recherche prospectifs:

- ◆ La prise en compte de décisions dépendantes permettrait d'une part d'inclure les ratios de production dans l'optimisation du système et d'autre part de tenir compte de certains types d'associations de flexibilités que nous avons écartés de notre étude.
- ◆ L'extension aux systèmes d'assemblage est également à considérer. Cette extension entraînerait alors des modifications sensibles au niveau de la modélisation et de l'analyse, en particulier en ce qui concerne l'en-cours et le système de transport.
- ◆ L'analyse des régimes transitoires entre des fonctionnements répétitifs doit être appréhendée dans la perspective d'une implantation réelle de la commande. En effet, dans une approche d'optimisation globale il faut également considérer le potentiel d'optimisation lié à cette problématique.

VI. Références bibliographiques

- [ALA 82] ALAIWAN, H. & MEMMI, G.
Algorithmes de recherche des solutions entières positives d'un système linéaire d'équations homogènes, Revue techn. Thomson-CSF, 1982, Vol. 14, n° 2, pp.125-135.
- [ALA 89] ALAIWAN, H. & TOUDIC, J.M.
Recherche des semi-flots, des verrous et des trappes dans les Réseaux de Petri, Technique et Science Informatiques, 1989, Vol. 4, n° 1, pp.103-112.
- [AMA 92] AMAR, S. CRAYE, E. & GENTINA, J.C.
A method of hierarchical specification and prototyping of FMS, IEEE-ETFA 92, Melbourne, Australia, 1992.
- [AMA 94] AMAR, S.
Systèmes automatisés et flexibles de production manufacturière: méthode de conception du système de coordination par prototypage orienté objet de la partie procédé, Thèse de Doctorat en Productique, Automatique et Informatique Industrielle, Université de Lille 1, 1994.
- [AUS 94] AUSFELDER, C. CASTELAIN, E. & GENTINA, J.C.
A method for hierarchical modeling of the command of Flexible Manufacturing Systems, IEEE Transactions on SMC, Vol 24, n° 4, pp.564-573.
- [BAC 89] BACELLI, F. & MAKOWSKI, A.
Queueing models for systems with synchronization constraints, Proceedings of the IEEE, Vol. 77, n° 1, January 1989, pp.138-160.
- [BAS 75] BASKETT, F. CHANDY, K.M. MUNTZ, R.R. & PALACIOS, F.G.
Open, closed and mixed networks of queues with different classes of customers, J.ACM, Vol. 22, n° 2, April 1975, pp.248-260.
- [BOI 91] BOIS, S.
Intégration de la gestion des modes de marche dans le pilotage d'un système automatisé de production, Thèse de Doctorat, Université de Lille 1, 1991.

- [BOU 88] BOUREY, J.P.
Structuration de la partie procédurale du système de commande de cellules de production flexibles dans l'industrie manufacturière, Thèse de Doctorat, Université de Lille 1, 1988.
- [BOU 93] BOUHCHOUCH, A. FREIN, Y. & DALLERY, Y.
Evaluation de performances de lignes de production bouclées par une méthode analytique, APII, Vol. 27, n° 3, pp.315-342.
- [CAM 90] CAMPOS, J.
Performance bounds for synchronized queueing networks, PhD thesis, Departamento de Ingenieria Electrica e Informatica, Universidad de Zaragoza, December 1990.
- [CAR 88] CARLIER, J. & CHRETIENNE, P.
Problèmes d'ordonnancement: modélisation / complexité / algorithmes, Editions Masson, PARIS, 1988.
- [CAS 87] CASTELAIN, E. CORBEEL, D. & GENTINA, J.C.
Apports de l'intelligence artificielle dans la réalisation d'un simulateur de Réseaux de Petri adaptatifs colorés, APII, Vol. 21, n° 4, 1987.
- [CER 88] CERNAULT, A.
La simulation des systèmes de production, Cepadues-Editions, Toulouse, 1988.
- [CHI 93] CHIOLA, G. ANGLANO, C. CAMPOS, J. COLOM, J.M. SILVA, M.
Operational analysis of timed Petri Nets and applications to the computation of performance bounds, Proceedings of 5th International Workshop on Petri Nets and Performance Models, PNPM 93, Toulouse, pp.128-137.
- [CHR 85] CHRETIENNE, P.
Analyse des régimes transitoire et asymptotique d'un graphe d'événements temporisé, Technique et Science Informatiques, Vol. 4, n° 1, 1985, pp.127-142.
- [COH 85] COHEN, G. DUBOIS, D. QUADRAT, J.P. & VIOT, M.
A linear system theoretic view of discrete-event processes and its use for performance evaluation in manufacturing, IEEE Transactions on Automatic Control, Vol. AC-30, n° 3, 1985, pp.210-220.

-
- [COH 89] COHEN, G. MOLLER, P. QUADRAT, J.P. & VIOT, M.
Algebraic tools for the performance evaluation of discrete event systems,
Proceedings of the IEEE, Vol. 77, 1989, pp.39-56.
- [COM 71] COMMONER, F. HOLT, A.W. EVEN, S. & PNUELI, A.
Marked directed graphs, Journal of Computer and System Sciences,
1971, Vol. 5, n° 5, pp.511-523.
- [CRA 89] CRAYE, E.
De la modélisation à l'implantation automatisée de la commande
hiérarchisée de cellules de production flexibles dans l'industrie
manufacturière, Thèse de Doctorat, Université de Lille 1, 1989.
- [CRU 91] CRUETTE, D.
Méthodologie de conception des systèmes complexes à événements
discrets: application à la conception et à la validation de la commande de
cellules flexibles de production dans l'industrie manufacturière,
Thèse de Doctorat, Université de Lille 1, 1991.
- [DAL 84] DALLERY, Y.
Une méthode analytique pour l'évaluation des performances d'un atelier
flexible, Thèse de Doctorat en Automatique, Institut National Poly-
technique de Grenoble, 1984.
- [DAU 92] DAUZERE-PERES, S.
Planification et ordonnancement de la production: une approche intégrée
cohérente, Thèse de Doctorat en Automatique, Université Paul Sabatier de
Toulouse, 1992.
- [DES 95] DESEL, J. & ESPARZA, J.
Free choice Petri Nets, Cambridge Tracts in Theoretical Computer
Science, Cambridge University Press, 1995.
- [DIA 87] DIAZ, M.
Applying Petri Net based models in the design of systems, in Concurrency
& Nets, K. Voss & G.Rozenberg, Springer Verlag Berlin, 1987, pp.435-
460.
- [DIC 93] DICESARE, F. HARHALAKIS, G. PROTH, J.M. SILVA, M. &
VERNADAT, F.B.
Practice of Petri Nets in manufacturing, Chapman & Hall, London, 1993.

- [DUB 83] DUBOIS, D. & STECKE, K.E.
Using Petri Nets to represent production processes, Proceedings of the 22nd Conference on Decision and Control, San Antonio, Texas, Dec. 1983, pp.1062-1067.
- [DUB 90] DUBOIS, D. & STECKE, K.E.
Dynamic analysis of repetitive decision-free discrete-event processes: applications to production systems, Annals of Operations Research, 1990, Vol. 26, pp.323-347.
- [ELK 93] ELKHATTABI, S.
Intégration de la surveillance de bas niveau dans la conception des systèmes à événements discrets: application aux systèmes de production flexibles, Thèse de Doctorat, Université de Lille 1, 1993.
- [ERS 82] ERSCHLER, J. LEVEQUE, D. & ROUBELLAT, F.
Periodic loading of Flexible Manufacturing Systems, IFIP Congress, APMS, Bordeaux, France, 1982, pp.327-339.
- [FDI 89] FDIDA, S. & PUJOLLE, G.
Modèles de systèmes et de réseaux, Performances et Files d'Attentes, Editions Eyrolles, Paris, 1989.
- [FLO 85] FLORIN, G. & NATKIN, S.
Les Réseaux de Petri stochastiques, Technique et Science Informatiques, 1985, Vol. 4, n° 1, pp.143-159.
- [FRE 88] FREIN, Y. DALLERY, Y. PIERRAT, J.J. & DAVID, R.
Optimisation du routage des pièces dans un atelier flexible par des méthodes analytiques, APII, 1988, Vol. 2, n° 5, pp.489-508.
- [GAU 90] GAUBERT, S.
An algebraic method for optimizing resources in timed event graphs, Lecture Notes in Control and Information, Springer Verlag, 1990, n° 144, pp.957-966.
- [GEN 88] GENTINA, J.C. BOUREY, J.P. & KAPUSTA, M.
Colored adaptive structured Petri Nets - a tool for the automatic synthesis of hierarchical control of Flexible Manufacturing Systems, Computer Integrated Manufacturing Systems, Butterworth Heinemann, February 1988, Vol. 1, n° 1, pp.39-47.

-
- [GOT 93] GOTHA
Les problèmes d'ordonnancement, Recherche opérationnelle, 1993,
Vol. 27, n° 1, pp.77-150.
- [HAM 91] HAMMADI, S.
Une méthode d'ordonnancement minimisant les temps d'attente et de
transit dans les systèmes de production flexible de type job-shop,
Thèse de Doctorat, Université de Lille 1, 1991.
- [HAN 89] HANEN, C.
Optimizing microprograms for recurrent loops on pipelined architectures
using timed Petri Nets, in Advances in Petri Nets, LNCS 424, 1989,
Springer Verlag, New York, pp.236-261.
- [HAR 92] HARHALAKIS, G. NAGI, R. & PROTH, J.M.
Hierarchical modeling approach for production planning, Proceedings of
the IFAC-INCOM 1992, pp.21-26.
- [HIL 87] HILLION, H.P. PROTH, J.M. & XIE, X.L.
A heuristic algorithm for the periodic scheduling and sequencing job-shop
problem, IEEE, 26th Conference on Decision and Control, Los Angeles,
1987, pp.612-617.
- [HIL 89a] HILLION, H.P. & PROTH, J.M.
Performance evaluation of job-shop systems using timed event-graphs,
IEEE Transactions on Automatic Control, 1989, Vol. 34, n° 1, pp.3-9.
- [HIL 89b] HILLION, H.
Modélisation et analyse des systèmes de production discrets par les
Réseaux de Petri temporisés, Thèse de Doctorat en Mathématiques,
Université Pierre et Marie Curie, Paris VI, 1989.
- [HO 83] HO, Y.C. & CASSANDRAS, C.
A new approach to the analysis of discrete event dynamic systems,
Automatica, Vol. 19, n° 2, 1983, pp.149-167.
- [HO 87] HO, Y.C.
Performance evaluation and perturbation analysis of discrete event
dynamic systems, IEEE Transactions on Automatic Control, Vol AC-32,
n° 7, July 1987.

- [HSU 93] HSU, L.F TAPIERO, C.S. & LIN, C.
Network of queues modeling in Flexible Manufacturing Systems:
a survey. Recherche opérationnelle / Operations research, AFCET-
Gauthiers-Villars, Vol. 27, n° 2, 1993, pp.201-248.
- [HUV 94] HUVENOIT, B.
De la conception à l'implantation de la commande modulaire et hiérarchisée
de systèmes flexibles de production manufacturière, Thèse de Doctorat en
Productique, Université de Lille 1, 1994.
- [HWA 86] HWANG, S.
Part selection problems in Flexible Manufacturing Systems planning stage,
Proceedings of the second ORSA TIMS Conference on Flexible Manu-
facturing Systems, Ann Arbor, Michigan, USA, Elsevier 1986, pp.297-
309.
- [KAP 88] KAPUSTA, M.
Génération assistée d'un graphe fonctionnel destiné à l'élaboration struc-
turée du modèle de la partie commande pour les cellules de production
flexibles dans l'industrie manufacturière, Thèse de Doctorat, Université de
Lille 1, 1988.
- [KAR 84] KARMARKAR, N.
A new polynomial time algorithm for linear programming.
Combinatorica, 1984, n° 4, pp.373-395.
- [KIM 83] KIMEMIA, J. & GERSHWIN, S.B.
An algorithm for the computer control of a Flexible Manufacturing
System, IIE Transactions, Vol. 15, n° 4, 1983, pp.353-362.
- [LAF 91] LAFTIT, S.
Graphes d'événements déterministes et stochastiques: application aux
systèmes de production, Thèse de Doctorat en Mathématiques, Paris IX
Dauphine, 1991.
- [LAF 92] LAFTIT, S. PROTH, J.M. & XIE, X.L.
Optimization of invariant criteria for event graphs, IEEE Transactions on
Automatic Control, 1992, Vol. 37, n° 5, pp.547-555 .
- [LON 93] LONG, J.
Sur la conduite hiérarchisée des systèmes flexibles de production,
Thèse de Doctorat en Productique, Institut National Polytechnique de
Grenoble, 1993.

- [MAR 79] MARIE, R.
An approximate analytical method for general queueing networks, IEEE Transactions on Software Engineering, 1979, Vol. SE-5, n° 5, pp.530-538.
- [MAR 82] MARTINEZ, J. & SILVA, M.
A simple and fast algorithm to obtain all invariants of a generalized Petri Net, in Application and Theory of Petri nets, Eds. C. Girault & W. Reisig, Springer Verlag, New York, 1982, pp.301-310.
- [MAR 89] MARSAN, A.
Stochastic Petri Nets: an elementary introduction, Lecture Notes in Computer Science, LNCS 424, Springer Verlag, 1989, Vol. 424, pp.1-29.
- [MAZ 89] MAZZOLA, J.B. NEEBE, A. & DUNN, C.V.R.
Production planning of a Flexible Manufacturing System in a material requirements environment, The International Journal of Flexible Manufacturing Systems, Kluwer Academic Publishers, Boston, 1989, Vol. 1, pp.115-142.
- [MOL 82] MOLLOY, M.K.
Performance analysis using stochastic Petri Nets, IEEE Transactions on Computers, 1982, Vol. 31, n° 9, pp.913-917.
- [MUN 91] MUNIER, A.
Résolution d'un problème d'ordonnancement cyclique à itérations indépendantes et contraintes de ressources, RAIRO, Recherche Opérationnelle, 1991, Vol. 25, n° 2, pp.161-182.
- [MUR 89] MURATA, T.
Petri Nets: properties, analysis and application, Proceedings of the IEEE, 1989, Vol. 77, n° 4, pp.541-580.
- [OHL 94a] OHL, H. CASTELAIN, E. & GENTINA, J.-C.
Synchrony theory applied to control problems in Flexible Manufacturing Systems, IEEE Conference on Systems, Man & Cybernetics, October 1994, San Antonio, Texas, Vol. 2, pp.1689-1694.
- [OHL 94b] OHL, H. CAMUS, H. CASTELAIN, E. & GENTINA, J.-C.
A heuristic algorithm for the computation of cyclic schedules and the necessary WIP to obtain optimal cycle time, Rensselaer's Fourth International Conference on Computer Integrated Manufacturing and Automation Technology, CIMAT 1994, Troy, New York, Vol. 1, pp.339-344.

- [OHL 95] OHL, H. CAMUS, H. CASTELAIN, E. & GENTINA, J.-C.
Petri Net modeling of ratio-driven Flexible Manufacturing Systems and implications on the WIP for cyclic schedules, accepted at the IEEE Conference on Systems, Man & Cybernetics, October 1995, Vancouver, Canada.
- [PRO 87] PROTH, J.M. & CAVAILLE, J.B.
Pratique de la simulation en production discontinue, Siprodis, Editions EC2, Collection Novotique, Paris, 1987.
- [PRO 95] PROTH, J.M. & XIE, X.L.
Les Réseaux de Petri pour la conception et la gestion des systèmes de production, Editions MASSON, 1995.
- [RAM 80] RAMAMOORTHY, C.V. & HO, G.S.
Performance evaluation of asynchronous concurrent systems using Petri Nets, IEEE Transactions on Software Engineering, Vol. SE-6, n° 5, 1980, pp.440-449.
- [SIF 80] SIFAKIS, J.
Performance evaluation of system using Nets, in Net theory and Applications, LNCS, Springer Verlag, Berlin, 1980, pp.307-319.
- [SIL 87] SILVA, M.
Towards a synchrony theory for p/t nets, in Concurrency and Nets, Eds K. Voss, Springer Verlag, Berlin, 1987, pp.435-460.
- [SIL 89] SILVA, M. & VALETTE, R.
Petri Nets and Flexible Manufacturing, in Advances in Petri Nets, LNCS 424, Springer Verlag, New York, 1989, pp.374-417.
- [SIL 92] SILVA, M. & MURATA, T.
B-Fairness and structural B-Fairness in Petri Net models of concurrent systems, Journal of Computer and System Science, Academic Press, Vol. 44, n° 3, pp.447-477.
- [SOL 79] SOLBERG, J.J.
Analytical performance evaluation of Flexible Manufacturing Systems, Proceedings IEEE, Conference on Decision and Control, San Diego 1979, pp.640-644.

- [STE 88] STECKE, K.E. & ILYONG, K.
A study of FMS part type selection approaches for short-term production planning, *The International Journal of Flexible Manufacturing Systems*, 1988, Vol. 1, n° 1, pp.7-29.
- [STE 91] STECKE, K.E. & KIM, I.
A flexible approach to part type selection in flexible flow systems using part mix ratios, *Int. J. Prod. Res.*, 1991, Vol. 29, n° 1, pp.53-75.
- [STE 92] STECKE, K.E.
Procedures to determine part mix ratios for independent demands in Flexible Manufacturing Systems, *IEEE Transactions on Engineering Management*, 1992, Vol. 39, n° 4, pp.359-369.
- [SUR 89] SURI, R.
Perturbation analysis: the state of the art and research issues explained via the GI/G/1 queue, *Proceedings of the IEEE*, Vol. 77, n° 1, January 1989.
- [TAW 95] TAWEGOUM, R.
Contrôle temps réel de déroulement des opérations dans les systèmes de production flexibles, Thèse de Doctorat, Université de Lille I - Ecole Centrale de Lille, Avril 1995.
- [TOG 92] TOGUYENI, A.
Surveillance et diagnostic en ligne dans les ateliers flexibles de l'industrie manufacturière, Thèse de Doctorat, Université de Lille 1, 1992.
- [TOU 82] TOUDIC, J.M.
Algorithmes d'algèbre linéaire pour l'analyse structurelle des Réseaux de Petri, *Revue technique Thomson-CSF*, Vol. 14, n° 1, pp.137-155, 1982.
- [VAL 86] VALETTE, R.
Nets in production systems, *Lecture Notes in Computer Science*, Springer Verlag, 1986, Vol. 255, pp.191-217.
- [XIE 89] XIE, X.L.
Real time scheduling and routing for Flexible Manufacturing Systems with unreliable machines, *RAIRO, Recherche Opérationnelle*, 1989, Vol. 23, n° 4, pp.355-374.
- [XIE 91] XIE, X.L.
Hierarchical production control of a Flexible Manufacturing System, *Applied Stochastic Models and Data Analysis*, 1991, Vol. 7, pp.343-360.

