HAL
open science

Transport automatisé dans les systèmes de fabrication de semi-conducteurs : nouvelles approches de gestion tactique et opérationnelle

Jairo Rafael Montoya Torres

To cite this version:

Jairo Rafael Montoya Torres. Transport automatisé dans les systèmes de fabrication de semi-conducteurs : nouvelles approches de gestion tactique et opérationnelle. Génie des procédés. Ecole Nationale Supérieure des Mines de Saint-Etienne, 2005. Français. NNT : . tel-00793340

HAL Id: tel-00793340

https://theses.hal.science/tel-00793340v1

Submitted on 22 Feb 2013

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

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

École Nationale Supérieure
des Mines de Saint-Étienne
N° d'ordre : 381GI

Université Jean Monnet
de Saint-Étienne

THÈSE

présentée par

Jairo Rafael Montoya Torres
Ingénieur, M.Sc.

en vue de l'obtention du titre de

DOCTEUR

de l'Ecole Nationale Supérieure des Mines de Saint-Etienne
et de l'Université Jean Monnet de Saint-Etienne

(spécialité : Génie Industriel)

TRANSPORT AUTOMATISÉ DANS LES SYSTÈMES DE FABRICATION DE SEMI-CONDUCTEURS : Nouvelles approches de gestion tactique et opérationnelle

soutenue publiquement à Gardanne le 29 novembre 2005

Composition du jury

Président et rapporteur	Prof. B. Grabot	École Nationale d'Ingénieurs de Tarbes
Rapporteurs	Dr.hdr. P. Lacomme	Université Blaise Pascal Clermont-Ferrand
	Prof. E. Yücesan	INSEAD, Fontainebleau
Examinateurs	Prof. J.P. Campagne	INSA de Lyon
	Prof. S. Dauzère-Pérès	École des Mines de Saint-Étienne
	Dr. H. Marian	École des Mines de Saint-Étienne
Invité	Ir. L. Vermariën	STMicroelectronics, Rousset

Thèse préparée dans le Centre Génie Industriel et Informatique et le Centre Microélectronique de Provence *Georges Charpak* de l'Ecole des Mines de Saint-Etienne

"Mind the gap"
(heard in the London Underground)

Remerciements

Les travaux présentés dans cette thèse ont été réalisés dans le cadre d'une Convention Industrielle de Formation par la Recherche (CIFRE n° 621/2002) accordée par l'Association Nationale de la Recherche Technique (ANRT), en collaboration avec STMicroelectronics. Ce travail fait aussi partie du projet "Rousset 2003-2008".
Je tiens à exprimer ma gratitude à mes directeurs de thèse. Toute ma gratitude envers Stéphane DAUZÈRE-PÉRÈS, complice de tous les jours qui a rendu ce travail moins solitaire. J'ai réellement apprécié et profité de toutes nos discussions, techniques et autres, ainsi que de ses conseils et de sa bonne humeur. J'espère que notre collaboration continuera encore longtemps. Mes remerciements à Jean-Pierre CAMPAGNE, pour sa disponibilité constante, et la confiance et le soutien qu'il m'a témoigné dès ma candidature à ce projet de collaboration académique et industrielle. Merci aussi à Hélène MARIAN qui, malgré des contraintes d'ordre privé, a toujours été disponible et a su s'investir dans le co-encadrement de mes travaux de thèse.
Mes remerciements vont ensuite aux autres membres du jury : Bernard GRABOT, Enver YÜCESAN et Philippe LACOMME pour avoir accepté d'être les rapporteurs de ma thèse malgré leur emploi du temps fort chargé, et pour leurs suggestions d'amélioration du manuscrit ; et Leon VERMARIËN, mon tuteur industriel, pour son soutien, sa disponibilité et ses réponses à toutes mes questions sur la microélectronique. Merci à Bernard pour avoir accepté la tâche de présider mon jury.
Presque 6 années passées en France (ou devrais-je dire en Europe ?) m'ont permis de me confronter à plusieurs différences culturelles qui m'ont énormément enrichi. Pendant ces années j'ai eu l'opportunité de rencontrer des personnes formidables, aussi bien au niveau professionnel que personnel. Il est difficile de les mentionner toutes sans courir le risque d'en oublier une grande partie... Je m'attacherais donc seulement à indiquer que j'ai bien apprécié les moments passés et l'apprentissage culturel avec mes camarades et mes amis et amies durant (et après) mon séjour à Grenoble, et tout particulièrement avec Igor et Mavi, Cozumel et Eric, Astrid, Alejandra, Claude, Tetsu et Carla, German et sa famille, Juan Pedro et Claudia et Asta, ainsi que les pauses-café (pauses-chocolat pour certains et pauses-thé pour d'autres) et les déjeuners sous le soleil provençal (même en hiver !...) dans le jardin de "la p'tite maison du Pesquier". Impossible d'oublier les repas fort agréables chez les ROUGEON et chez les FOURNIER : ils m'ont permis de mieux vivre, comprendre et appréhender la culture et la cuisine françaises.
Je voudrais aussi remercier Gloria L. RODRÍGUEZ V. pour sa collaboration dans le développement informatique des modèles de la fab ; Julie CHAPRON et Sana BELMOKHTAR, pour l'amitié et les encouragements qu'elles m'ont témoignés et les discussions et débats très animés en dehors du cadre de travail, Cathy WOLOSEWICZ pour avoir su me supporter dans le bureau et pour la relecture de ce manuscrit. Je tiens à vivement remercier Carolina DEL VALLE, pour son amitié inconditionnelle et son soutien constant depuis toujours ; Roberto Enrique MONTOYA V., pour ses précieux conseils et son soutien ; Sandra CARBONELL, pour avoir su me comprendre et pour son soutien et ses conseils assertifs ; et María Cecilia FORERO M. qui avec sa sagesse et sa richesse intérieure m'a toujours inspiré pour faire de mon mieux.
Je ne saurais commencer ce mémoire de thèse sans remercier du fond de mon cœur mon père, ma mère et ma sœur, pour leurs conseils et leur soutien tout au long de mes études et pendant mon séjour en France.

Jairo R. MONTOYA T.
Marseille, 2005

Table des matières

Remerciements **v**

1 Introduction générale **1**
1.1 L'industrie des semi-conducteurs 1
1.2 Objectif de la thèse et aperçu des contributions 3
1.3 Plan de lecture 5

I Contexte, modélisation et analyse de la problématique **7**

2 La fabrication de wafers **9**
2.1 Introduction 9
2.2 Le processus de fabrication 12
2.3 L'agencement des machines dans la fab 14
2.4 Le transport des wafers dans la fab 15
2.5 Mesure de la performance de la fabrication 16
2.5.1 Temps de cycle 16
2.5.2 Niveau d'encours 17
2.5.3 Débit du système (throughput rate) 18
2.6 Difficultés pour la gestion des opérations 19
2.7 Conclusion 20

3 Le transport automatisé de wafers **21**
3.1 Notions préliminaires 21
3.2 Conception de l'AMHS pour les fabs 22
3.2.1 Système de transport inter-bay 22
3.2.2 Système de transport intra-bay 24
3.2.3 Taille de la flotte de véhicules 24
3.3 Le pilotage du système de transport 25
3.3.1 Le problème de dispatching 25
3.3.2 Le problème de routage 27
3.3.3 Le positionnement des véhicules 27

3.3.4 La recherche de tâches 28
3.4 Travaux sur la gestion des véhicules dans les fabs 29
3.5 Synthèse et démarche 30

4 Modélisation conceptuelle de la fab 33
4.1 Modélisation des systèmes industriels 33
4.2 Modèle du système de fabrication 35
4.2.1 Méthodologie : La “Process-Interaction Approach” 35
4.2.2 Modèle du système physique de production 35
4.2.3 Modèle du système de supervision de la fab 37
4.3 Modélisation du processus de fabrication 39
4.3.1 Une approche par les réseaux de Petri 39
4.3.2 Le modèle 40
4.4 Conclusion 45

5 Présentation et analyse du cas industriel 47
5.1 Présentation du cas industriel 47
5.1.1 Le processus de fabrication et le système de transport 47
5.1.2 Particularités du cas industriel 49
5.2 Etude de simulation 50
5.2.1 Présentation succincte du modèle 50
5.2.2 Règles de pilotage du système de transport 51
5.2.3 Plan d’expériences 53
5.2.4 Résultats 54
5.3 Conclusion 59

II Optimisation de la gestion du système de transport automatisé 61

6 Définition tactique des zones 65
6.1 Problème tactique de positionnement des véhicules 65
6.2 Formulations mathématiques 66
6.2.1 Considérations pour la modélisation et hypothèses 66
6.2.2 Programmes linéaires en nombre entiers (PLNE) 67
6.2.2.1 Minimisation du temps maximal de couverture 67
6.2.2.2 Prise en compte du taux d’utilisation des véhicules 74
6.2.2.3 Prise en compte de la demande des zones 77
6.3 Extensions des modèles 81
6.4 Une méthode heuristique 82
6.4.1 Algorithme Glouton 82
6.4.2 Résultats numériques 82
6.5 Conclusion 85

7 Gestion opérationnelle des véhicules 87
7.1 Rappel de la problématique opérationnelle 87
7.2 Politiques de gestion des zones et des véhicules 88
7.2.1 Politique ZVBRet : Définition de zones, nombre de véhicules borné par zone/retour du véhicule 91
7.2.2 Politique ZVBProp : Définition de zones, nombre de véhicules maximum borné par zone, ré-affectation selon le nombre actuel de véhicules 92
7.2.3 Politique ZVBDem : Définition de zones, nombre de véhicules maximum borné par zone, ré-affectation selon la demande dans les zones 94
7.2.4 Politique ZVBDemFab : Définition de zones, nombre de véhicules maximum borné par zone, ré-affectation selon la demande dans la fab 95
7.3 Couplage optimisation-simulation 96
7.3.1 Mise en place de la solution de la planification tactique 97
7.3.2 Etude sur le changement des paramètres tactiques 101
7.3.2.1 Changement du nombre des véhicules affectés aux zones . . 101
7.3.2.2 Ajout d'une borne pour le temps de couverture 102
7.4 Sur l'impact de véhicules partagés 104
7.5 Conclusion 105

8 Conclusion générale et perspectives 107

Bibliographie 119

Liste des tableaux

2.1 Description des étapes de la fabrication de wafers 11

3.1 Stratégies de pilotage des véhicules 29

5.1 Facteurs et niveaux du plan d'expériences 54
5.2 Résultats des simulations 56
5.3 Temps moyen de livraison 58

6.1 Exemple du problème de symétrie 70
6.2 Sensibilité du PLNE par rapport aux points dans J 71
6.3 Temps de résolution du modèle **VP1** (en sec., sans coupes) 72
6.4 Impact des coupes sur les temps de calcul (en sec.) 72
6.5 Résultats du modèle **VP1** 73
6.6 Impact des coupes sur les temps de calcul (en sec.) pour le modèle **VP2** . . 76
6.7 Résultats du modèle **VP2** 77
6.8 Impact des coupes sur les temps de calcul (en sec.) pour le modèle **VP3** . . 79
6.9 Résultats du modèle **VP3** 80
6.10 Résultats de l'algorithme glouton 84

7.1 Comparaison des politiques opérationnelles 89
7.2 Résultats des simulations : Politique ZVBRet 92
7.3 Résultats des simulations : Politique ZVBProp 93
7.4 Résultats des simulations : Politique ZVBDem 95
7.5 Résultats des simulations : Politique ZVBDemFab 96
7.6 Facteurs et niveaux du plan d'expériences 97
7.7 Résultats des simulations de l'implémentation de la planification tactique : Règle STD, 1 zone (scénario de référence) et 5 zones 99
7.8 Résultats des simulations de l'implémentation de la planification tactique : Règle STD, 10 et 15 zones 99
7.9 Résultats des simulations de l'implémentation de la planification tactique : Règle CL-FCFS, 1 zone (scénario de référence) et 5 zones 100
7.10 Résultats des simulations de l'implémentation de la planification tactique : Règle CL-FCFS, 10 et 15 zones 100

7.11 Distribution des véhicules dans les zones 101
7.12 Résultats de la modification de la distribution des véhicules dans les zones . 102
7.13 Résultats des simulations avec une borne pour le temps de couveture 103
7.14 Résultats des simulations : Politique ZVBDemFab avec véhicules partagés . 105

Table des figures

1.1 La loi de Moore 2

2.1 Passage du wafer vers circuit intégré [120] 10
2.2 Lingot de silicium et plaquettes vierges 11
2.3 Séquence des étapes de la fabrication des wafers [120] 12
2.4 Image en microscopie électronique d'un composant possédant trois niveaux de métal [102] 13
2.5 Concepts de base de la photo-lithographie [120] 13
2.6 Les procédés de gravure et strippage [120] 14
2.7 Configurations possibles d'une fab multi-bay 15
2.8 Encours vs. temps de cycle et throughput 18
2.9 Relation entre temps de cycle et throughput 19

3.1 Configurations *spine* et en périmètre 23
3.2 Pilotage opérationnel d'un système automatisé de transport 26

4.1 Environnement de l'unité de fabrication 36
4.2 Modèle du système de fabrication (*Fab*) 36
4.3 Modèle *FabResources* 37
4.4 Modèle *FabTransporters* 37
4.5 Modèle du système de supervision (*FabController*) 38
4.6 Notation graphique du RdPCT 40
4.7 Modèle général du processus de fabrication 41
4.8 Modèle du transport des lots 42
4.9 Modèle de la fabrication de lots 44

5.1 Configuration du réseau de transport (AMHS) 49
5.2 Temps de cycle et utilisation des véhicules, règle FCFS 56
5.3 Temps de cycle et utilisation des véhicules, règle STD 57
5.4 Temps de cycle et utilisation des véhicules, règle CL-FCFS 57
5.5 Temps de transport 58

6.1 Exemple de découpage du réseau de transport en zones 66

6.2 Description de l'algorithme GLOUTON 83
6.3 Temps maximal de couverture : PLNE *versus* Glouton 84

7.1 Comparaison des résultats : Temps de cycle et utilisation des véhicules . . . 98

8.1 Intégration de notre démarche pour l'optimisation, la simulation et l'aide à la décision 109

Chapitre 1

Introduction générale

1.1 L'industrie des semi-conducteurs

Tout commença il y a une cinquantaine d'années avec l'invention du premier transistor qui donna naissance à l'industrie de la microélectronique [35]. Bell Telephone Laboratories (Bell Labs) voulait développer des systèmes électroniques plus fiables que les dispositifs électromécaniques. En 1947, Bell Labs produisirent le premier transistor. En 1959, Fairchild Semiconductor Corp. décrivit le processus de fabrication d'un transistor, au même moment où Texas Instruments présenta des dispositifs électroniques et leurs interconnexions dans une seule pièce d'un matériau semi-conducteur. Les deux inventions constituent les bases de la fabrication des circuits intégrés que nous connaissons aujourd'hui. Quelques années plus tard, en 1960, Texas Instruments fabriqua le premier circuit intégré commercialisable. Même si les technologies de fabrication des composants électroniques ont beaucoup évolué depuis l'invention du premier transistor, certains principes ont été conservés. Bien entendu, la technologie est subordonnée aux contraintes qui résultent des propriétés fondamentales des semi-conducteurs.

Depuis sa naissance, l'industrie des semi-conducteurs s'est très rapidement développée jusqu'à devenir aujourd'hui l'un des principaux moteurs de l'économie mondiale. En effet, on trouve des composants microélectroniques à base de semi-conducteurs dans presque tous les produits de la vie courante : téléphonie fixe et mobile, informatique, électroménager, automobile,... De plus, ses clients reçoivent, pour le même coût, des produits de plus en plus performants, une tendance établie par la loi de Moore[1]. La figure 1.1 présente deux exemples représentatifs de cette tendance du secteur de la microélectronique : les microprocesseurs et les mémoires. Pour les deux produits, nous pouvons observer qu'au cours des années, le nombre de transistors par puce augmente alors que leur taille diminue. Pour les microprocesseurs, la progression est un peu plus faible car les lois d'échelle ne sont pas aussi facilement transposables que dans le cas des mémoires. Cette loi dépend de deux facteurs : d'une part le développement d'une technologie permettant de "graver" des composants de plus en plus petits, et d'autre part le développement d'applications nécessitant des circuits

[1] La loi de Moore est une observation historique réalisée par Gordon Moore, co-fondateur d'Intel, qui dit que la demande du marché (et la réponse de l'industrie de semi-conducteurs) exprimée en nombre de fonctionnalités par puce (octets, transistors) double tous les 1,5 à 2 ans [93, 94]. Aussi vue comme une prophétie, la loi de Moore a été une macro-tendance et un indicateur clé pour le développement des produits dans cette industrie pendant les 30 dernières années.

toujours plus petits, performants, rapides, fiables et peu coûteux. Dans la figure 1.1, l'axe de droite montre la progression des technologies de lithographie optique.

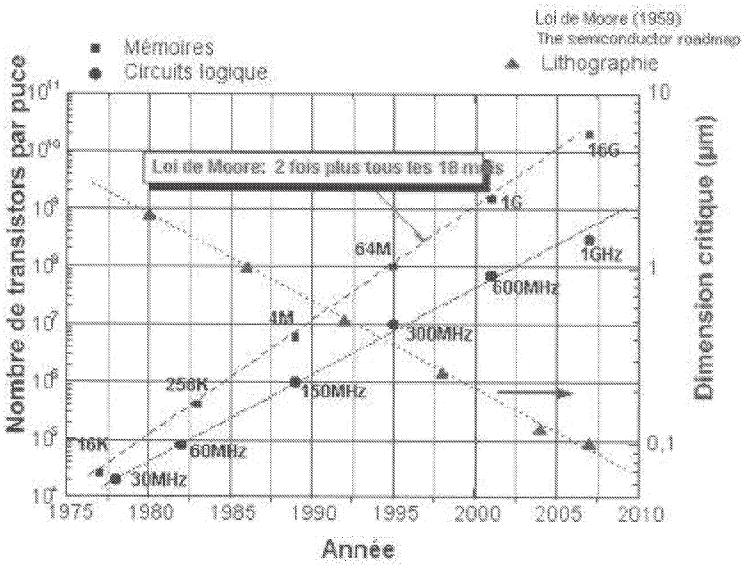


FIG. 1.1 – La loi de Moore

Afin de maintenir ce niveau de croissance, l'industrie de semi-conducteurs a focalisé ses efforts sur l'amélioration des opérations clés de la fabrication. Aujourd'hui, ce secteur industriel est caractérisé à la fois par un rythme accéléré de renouvellement des technologies et par une complexité croissante des procédés aussi bien au niveau des contraintes de production que du nombre d'opérations élémentaires. De plus, afin de réduire les coûts totaux associés à la fabrication, la taille des plaquettes augmente aussi. En passant de 200mm vers 300mm, les équipement sont de plus en plus sophistiqués avec un niveau d'automatisation de plus en plus élevé.
L'arrivée de la nouvelle génération de wafers de 300mm de diamètre est désormais une réalité pour les fabricants de semi-conducteurs. C'est une grande opportunité pour l'industrie de réduire les coûts de fabrication et, aussi, d'optimiser la performance industrielle à travers un effort collectif et systématique entre les fournisseurs, les fabricants et les clients (i.e. le *Global Joint Guidance for 300mm Semiconductor Factories*[2]).
Aujourd'hui, l'industrie de semi-conducteurs a besoin de nouvelles technologies et de recherches approfondies dans tous les domaines, comme suggéré par l'*International Technology Roadmap for Semiconductors* (ITRS)[3] [41]. Les besoins en performance imposés par

[2]Cette alliance est formée par deux consortia : I300I et J300. Le I300I est l'*International 300mm Initiative* et est formée par les fabricants des Etats-Unis d'Amérique, Europe, Taiwan et Corée, tandis que J300 est un consortium formé par les fournisseurs et fabricants du Japon. Les deux groupes ont été très actifs dans les discussions et le développement de standards relatifs à la fabrication de wafers de 300mm de diamètre.

[3]Dans l'industrie des semi-conducteurs, traditionnellement, les lignes directrices pour le développement de la recherche et les nouvelles technologies sont dictées par l'ITRS. Ce document est une source d'information extrêmement riche pour les fournisseurs, fabricants et chercheurs associés à ce secteur industriel. L'ITRS est périodiquement publié (1992, 1994, 1997, 1999, 2001, 2003) et est sponsorisé par la SIA (Se-

la demande ont été des lignes directrices pour le développement de nouveaux produits et de nouvelles technologies.
Mais l'arrivée des wafers de 300mm génère aussi des contraintes ergonomiques [122] qui exigent l'automatisation de toute la manutention et du transport des wafers dans l'unité de fabrication ("fab"). En effet, une cassette[4] contenant un lot de wafers de 300mm est trop lourde (environ 12 kg) pour être transportée par des opérateurs sans courir le risque de causer, à terme, des traumatismes musculaires permanents [122]. Une manipulation manuelle systématique pourrait aussi endommager le produit. Un système automatisé de transport (*Automated Material Handling System*, AMHS) est donc nécessaire pour augmenter la sécurité du personnel et du produit. De plus, l'automatisation du transport interne de la fab a aussi une importance économico-financière. Avec le passage de 200mm à 300mm de diamètre, la surface des plaquettes a augmentée de 50%, ce qui veut dire qu'à taille de puce constante, le coût d'une plaquette en cours de fabrication a augmenté dans les mêmes proportions. Il est donc indispensable d'avoir un système automatique de gestion permettant un suivi plus sûr et plus précis des wafers dans la fab.

1.2 Objectif de la thèse et aperçu des contributions

Dans cette thèse, nous nous intéressons au système automatisé de transport de wafers dans les fabs. Les travaux précédents dans la littérature des semi-conducteurs se sont focalisés sur les règles de priorité (dispatching) pour servir les demandes de transport (sélection d'une demande et sélection d'un véhicule). Les travaux publiés considèrent des modèles de simulation *ad hoc* pour tester l'efficacité de certaines règles de dispatching de véhicules, sans une véritable démarche d'optimisation du système. En outre, le problème de sélection de requêtes de transport a aussi été étudié dans ces travaux concernant d'autres systèmes de fabrication automatisée (e.g. FMS, jobshops) avec, en plus, un intérêt pour la gestion du trafic dans le réseau (routage des véhicules et prévention des collisions). Dans cette thèse, nous nous intéressons à une autre analyse du transport des wafers. En raison de la taille du réseau et de sa configuration, nous étudions le problème à partir d'un autre point de vue. Notre objectif est d'analyser le comportement de l'AMHS comme une ressource au service de la production. Lorsque les règles de dispatching, de routage et de gestion du trafic sont définies, notre problème est de gérer les véhicules efficacement dans le système de façon à optimiser le temps de service des demandes, tout en utilisant le moins de véhicules possibles : le nombre de véhicules étant une ressource limitée, l'objectif est de les distribuer au mieux dans le réseau et de mettre en place des règles d'affectation pertinentes. Notre but est donc de nous assurer de la présence de véhicules aux endroits où il y aura des requêtes. Bien évidement, il existe une interaction entre le dispatching des véhicules et la stratégie de service. Nous explorons également ce point dans cette thèse.

Dans ce travail, nous ne nous limitons pas simplement proposer un modèle de simula-

miconductor Industry Association), EECA (European Electronic Component Association), JEITA (Japan Electronics & Information Technology Industries Association), KSIA (Korean Semiconductor Industry Association) et TSIA (Taiwan Semiconductor Industry Association). Dans cette thèse, nous ferons référence à la dernière édition du rapport ITRS publiée en 2003 lorsque cela sera nécessaire.

[4]Pour la fabrication de wafers de 300mm de diamètre, la cassette standard utilisée par les fabricants est appelée FOUP (*Front Opening Unified Pod*). Cette cassette est une boîte hermétique dans laquelle sont contenus les wafers lorsqu'ils ne sont pas en train d'être exécutés sur une machine. Elle les protège d'éventuelles particules polluantes dans la salle blanche.

tion pour notre système, nous proposons une démarche d'analyse et d'optimisation du transport de wafers en couplant des techniques classiques de la Recherche Opérationnelle (i.e. programmation linéaire en nombre entiers, méthodes heuristiques) avec un modèle de simulation à événements discrets. L'étape d'optimisation permet de trouver les valeurs optimales des paramètres du modèle de simulation, qui sera ensuite utilisé pour l'analyse du comportement du système en prenant en compte son évolution dynamique dans le temps. Les interactions entre la gestion du transport dans la fab et la gestion de la production sont importantes [61]. Cependant, cette recherche ne s'intéresse pas à l'ordonnancement de la production ni à l'étude des dites interactions. Nous sommes conscients de l'impact que la gestion de la logistique interne de la fab peut avoir sur la performance des stratégies de contrôle de la production. Nous implémentons donc dans notre modèle de simulation de l'AMHS des règles proches de la réalité pour la gestion des machines et des encours. Ces aspects seront discutés plus amplement dans le mémoire.

Ce travail de thèse étant réalisé dans le cadre d'un contrat CIFRE. La démarche proposée est mise en œuvre sur une nouvelle unité de fabrication de wafers de 300mm de diamètre. Notre démarche d'analyse et d'optimisation est cependant applicable à d'autres configurations de systèmes de fabrication, que ce soit une autre fab (de l'ancienne génération de 200mm, de 300mm ou même des générations à venir comme par exemple de 450mm) ou un système de production automatisée autre que les semi-conducteurs, car elle n'est pas liée à la configuration du réseau (*layout*) ni aux caratéristiques spécifiques des équipements dans le système. Aussi, puisque notre étude du système de transport considère celui-ci comme étant une ressource au service de la production qui doit être gérée de façon efficace, notre démarche peut être appliquée quelle que soit la politique de gestion de production implémentée.

Du point de vue de l'application industrielle, notre travail contribue à une compréhension plus ample du système de transport, de son fonctionnement, et de ses interactions avec les équipements de production. Cette contribution est mise en évidence notamment à travers la modélisation conceptuelle de l'unité de fabrication, de la construction du modèle (programme) de simulation lui-même, et de la mise en place d'une démarche d'optimisation. De plus, le couplage de notre démarche d'optimisation avec le modèle de simulation de la fab montre son intérêt pratique, notamment à un niveau tactique de prise de décisions. En effet, la cohérence des résultats du couplage optimisation-simulation nous permet de réaliser, dans la pratique, une analyse tactique préliminaire lors de la définition des éléments essentiels pour la configuration du réseau de transport (e.g. localisation des machines, définition du nombre de véhicules). Plusieurs possibilités peuvent être testées et comparées par rapport à des critères de performance choisis (e.g. temps de service, taux d'utilisation des véhicules). L'impact de ces décisions tactiques pourrait bien plus tard être validé dans une étape d'analyse opérationnelle à l'aide des modèles de simulation de la fab. Les coûts associés à l'implémentation des modèles de simulation pourront ainsi être diminués, en particulier parce que seuls les scénarii donnant les meilleures performances tactiques devront être considérés pour la simulation. Nous pouvons ainsi réduire le coût lié à la construction (ou mise à jours) des modèles de simulation. De plus, le temps dédié aux tests est aussi réduit, notamment lorsque les temps de calcul sont très importants (les modèles de la fab étant très détaillés, la simplification du niveau tactique constitue donc un grand atout pour l'analyse). Nous explorons ces bénéfices de notre contribution plus amplement dans le mémoire.

1.3 Plan de lecture

Cette thèse est divisée en deux grandes parties : “Contexte, modélisation et analyse de la problématique” et “Optimisation de la gestion du système de transport automatisé”. Nous recommandons une lecture linéaire afin d’appréhender les motivations et les objectifs de chaque chapitre et ainsi mieux comprendre les concepts présentés.

La première partie est composée de 4 chapitres et correspond à la description du contexte industriel de cette recherche et à la présentation et analyse d’un cas industriel. Notre positionnement dans la littérature scientifique est également présenté dans cette partie du mémoire.
Nous commençons par une présentation au chapitre 2 des principales notions sur la fabrication de wafers semi-conducteurs. Une telle présentation est essentielle pour une bonne compréhension des caractéristiques et de la complexité intrinsèque de ce secteur industriel. Après une description des principales étapes de la fabrication, ce chapitre s’intéresse aux indicateurs clés de la performance industrielle, ainsi qu’aux facteurs contribuant à la complexité de la gestion des opérations (logistiques et de production).
Le chapitre 3 est dédié à l’état de l’art sur la conception et la gestion du système de transport automatisé dans les unités de production de semi-conducteurs. Nous présentons tout d’abord quelques notions sur les systèmes automatisés pour le transport de produits dans la fab, qui seront exploités plus en détail dans ce mémoire. Ensuite, nous analysons la littérature scientifique sur le pilotage de tels systèmes et identifions ainsi ses manques.
Afin de mieux comprendre les relations entre les composants de l’unité de fabrication, le chapitre 4 est dédié à la modélisation conceptuelle de la fab. Nous utilisons des formalismes de modélisation pour décrire le système de production et le processus de fabrication, en intégrant les aspects production et transport. Ce cadre de modélisation permet également d’identifier les entités du système et leurs interactions afin de mettre en place une démarche de simulation et d’optimisation.
Notre travail de recherche étant réalisé en partenariat avec une compagnie de l’industrie de semi-conducteur, le chapitre 5 présente le cas industriel support de cette thèse. Dans une première partie, nous décrivons le processus de fabrication et la configuration du réseau de l’AMHS considéré dans cette recherche. Ensuite, nous développons une étude de simulation de différentes règles de gestion du système de transport (dispatching de véhicules) appliquées au cas industriel. Le principal objectif est d’identifier les paramètres critiques et ainsi, quantifier l’impact de la mise en place de certaines politiques de gestion du système. Les resultats de cette étude serviront de base pour la démarche d’optimisation développée dans la seconde partie de cette thèse.

A partir des résultats obtenus lors de l’étape de modélisation et d’analyse de l’unité de production, nous pouvons entamer une démarche d’optimisation. La seconde partie de cette thèse est divisée en 2 chapitres. Son objectif est d’améliorer la performance industrielle en mettant en place une démarche formelle pour la gestion et le pilotage du système de transport automatisé. Nous abordons ce problème à partir de deux niveaux de gestion : le niveau tactique et le niveau opérationnel.
Au chapitre 6, nous étudions le problème de distribution des véhicules dans le réseau à partir d’un point de vue tactique. Tout d’abord, le problème est formalisé à l’aide de programmes linéaires en nombres entiers en considérant les paramètres et contraintes pertinents à l’optimisation de la performance de la fab. En raison de la taille du réseau, une

résolution efficace en vue d'une démarche itérative d'aide à la décision n'est possible qu'à travers des méthodes heuristiques.
Les résultats obtenus au niveau tactique doivent ensuite être implémentés au niveau opérationnel. Le chapitre 7 présente quelques-unes des politiques pouvant être mises en place dans notre modèle de simulation. Un couplage entre les solutions de l'optimisation tactique et le modèle de simulation de la fab nous permettra de voir les effets de chacune de ces politiques sur la performance global du système.

Les conclusions de ce travail ainsi que des perspectives de recherche sont présentées à la fin du mémoire.

Première partie

Contexte, modélisation et analyse de la problématique

Chapitre 2

La fabrication de wafers

La fabrication des composants électroniques, que l'on trouve sur les cartes à puces, les cartes mémoires et les constituants de la microinformatique (microprocesseurs...), est réalisée aujourd'hui sur des substrats de silicium (wafers) de 200 et 300 mm de diamètre. Ce processus de fabrication est très complexe en termes de gestion des opérations et de la production. En effet, le procédé de fabrication consiste en une longue succession d'étapes très différentes. Ces étapes peuvent être schématisées par une succession d'empilements de matériaux en couches minces sur un support de silicium. Certains matériaux sont présents de façon très localisée, alors que d'autres sont en film quasiment continus. Chaque matériau a une application précise : diélectrique pour l'isolation, conducteurs pour les interconnexions, barrières de diffusion, passivation,...
Même si les technologies de fabrication des composants électroniques ont beaucoup évolué depuis l'invention du premier transistor, certains principes ont été conservés. Bien entendu, la technologie est subordonnée aux contraintes qui résultent des propriétés fondamentales des semi-conducteurs. Dans les paragraphes qui suivent, nous allons brièvement décrire les différentes étapes du processus de fabrication des wafers. Les conditions de fabrication peuvent varier d'une compagnie à une autre car les procédés utilisés par chaque fabricant dépendent du produit fabriqué et de la technologie utilisée. La description présentée ci-après est donc générique. Le lecteur intéressé par une description plus étendue et détaillée de la fabrication de wafers et des circuits intégrés peut se référer à [35, 129].

2.1 Introduction

Par définition, un circuit intégré (*Integrated Circuit*, IC) contient de millions de composants électroniques de taille microscopique. Le silicium est le matériau semi-conducteur le plus utilisé grâce à ses propriétés électriques. Ces composants sont formés à partir de l'empilement de couches de matériaux différents formant de régions isolées avec des caractéristiques électriques distinctes. Les composants sont inter-connectés afin de créer la fonctionnalité électronique souhaitée. Le produit final est une "puce" (ou *chip* en langue anglaise) très fine (entre 0,5 mm et 150 μm), rectangulaire et plate de 25 à 400 mm^2 de matériau semi-conducteur. Cette puce sera ensuite montée sur un support, encapsulée et soudée sur un PCB (*Printed Circuit Board*) et utilisée pour des applications en informatique, électroménager, audiovisuel, communication, etc.

La fabrication d'une puce de silicium est réalisée en trois grandes étapes : l'obtention du monocristal de silicium, la fabrication des wafers, et le packaging. La figure 2.1 schématise le passage du wafer vers le circuit intégré.

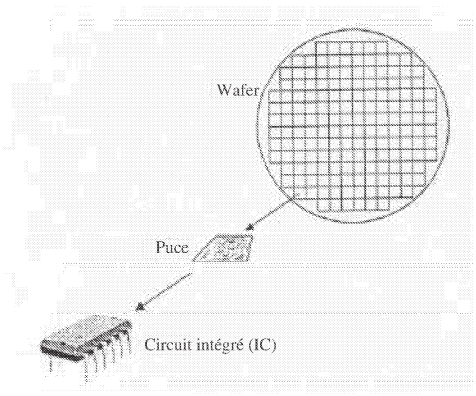


FIG. 2.1 – Passage du wafer vers circuit intégré [120]

Obtention du monocristal de silicium

Les matériaux semi-conducteurs sont la base de l'industrie micro-électronique. Le silicium est le matériau le plus utilisé, 95% de circuits intégrés (*chips*) sont fabriqués sur des substrats de silicium. Ce choix s'explique principalement par les excellentes propriétés et le bas coût du matériau. Typiquement le silicium de qualité métallurgique est obtenu à partir du sable. Ce silicium pur à 98-99% environ est impropre à l'utilisation pour l'industrie électronique qui nécessite au moins 99,9999% de pureté. Cette purification est obtenue en exploitant le procédé dit de purification physique par zone fondue. Afin d'obtenir une distribution uniforme des propriétés sur une grande surface du matériau semi-conducteur, le substrat de silicium doit être de structure mono-cristalline. Le procédé de Czochralski [124] est le plus utilisé pour obtenir cette structure. C'est seulement après que le lingot de silicium est découpé en tranches ou plaquettes de 100 μm à 1 mm d'épaisseur (voir figure 2.2), avec un diamètre de 200, 300 ou 450 mm. A partir de cet instant, il est indispensable d'opérer en atmosphère totalement dépoussiérée (salle blanche[1]).

La fabrication des wafers

La fabrication d'une puce sur une plaquette de silicium consiste en plusieurs types de procédés, lesquels sont répétés plusieurs fois. Leur but est d'ajouter, de modifier ou d'enlever des couches de matériaux sur certaines régions de la surface du wafer. Les procédés qui ajoutent des couches sont la déposition (chimique ou physique) et l'oxydation thermique.

[1] La salle blanche est un environnement contrôlé où sont fabriqués les circuits intégrés. Elle est spécialement conçue pour contrôler le flux d'air, la température et le niveau d'humidité de telle façon que le niveau de contamination en particules par unité de volume soit toujours en dessous d'un seuil préalablement défini. La classe de la salle blanche indique le nombre de particules de taille supérieure au micron par unité de volume d'air qui sont tolérables. La classe 100 indique qu'il n'y a pas plus de 100 particules par cm^3, et la classe 1 indique qu'il ne peut y avoir pas plus d'une particule de taille supérieure au micron par cm^3.

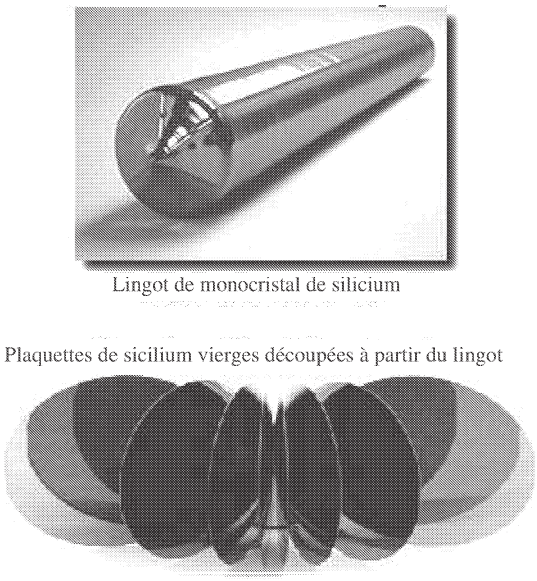


FIG. 2.2 – Lingot de silicium et plaquettes vierges

Les procédés qui modifient les couches sont la diffusion et l'implantation ionique. Le procédé de gravure est utilisé pour enlever du matériau. Un procédé appelé photo-lithographie est appliqué pour différencier les régions qui seront concernées par l'un ou l'autre des procédés. Le tableau 2.1 résume les principaux procédés et leur description. La section 2.2 décrit plus en détail ces différentes étapes.

TAB. 2.1 – Description des étapes de la fabrication de wafers

Processus	**Description**
Photo-lithographie	Transpose une structure sur la plaquette
Gravure (sèche ou humide)	Enlève du matériau de la surface de la plaquette
Implantation ionique	Définit des zones actives
Diffusion	Diffuse du matériau sur la surface de la plaquette
Oxydation	Ajout d'une couche de dioxyde de silicium
Déposition	Dépose d'un matériau diélectrique ou métallique
Polissage (CMP)	Diminue l'épaisseur de la plaquette

Le packaging

Une fois le wafer termine toutes les opérations de sa gamme, seules les puces de bonne qualité sont utilisables par la suite. Un wafer comportant de nombreux circuits identiques doit être découpé, puis chaque puce élémentaire sera montée sur un support (grille multi-patte) et enfin, après soudure des liaisons composant-pattes de sortie, le composant sera encapsulé (packaging). Le produit fini doit ensuite passer par une étape d'inspection qui mesure les caractéristiques de ses performances sous différentes contraintes. Une fois ce test réussi, le circuit intégrés peut être envoyé au client.

Dans cette thèse, nous nous intéressons à l'étape de fabrication des wafers semi-conducteurs. Comme nous avons vu, cette étape commence avec les plaquettes de silicium vierges et se termine juste avant le test et le découpage de chaque wafer en puces.

2.2 Le processus de fabrication

Cette section se focalise sur la séquence de procédés nécessaires pour la fabrication du circuit intégré. Chaque couche formée sur la plaquette a une fonctionnalité bien précise. Par conséquent, la séquence d'étapes nécessaires pour la fabrication de chaque couche doit être bien définie. Cette séquence s'appelle gamme opératoire (ou *process flow*, en langue anglaise). La figure 2.3 présente un schéma illustrant le caractère cyclique (re-entrant) des opérations de la gamme. Dans le langage des semi-conducteurs, le processus de fabrication est divisé en deux grandes parties : le FEOL (*Front-End Of the Line*) et le BEOL (*Back-End Of the Line*). Les étapes du FEOL se concentrent sur la fabrication des composants électroniques, alors que les étapes du BEOL sont dédiées à la mise en place des interconnections entre les composants électroniques. A l'exception du procédé de photo-lithographie, les opérations du FEOL et du BEOL sont complètement différentes. Le procédé de lithographie est nécessaire aussi bien en FEOL qu'en BEOL. La figure 2.4 est une image en coupe d'un wafer avec trois niveaux de métal qui montre à quel point les matériaux sont imbriqués. Comme nous avons déjà indiqué, certains matériaux sont présents de façon très localisée, comme les via en tungstène, alors que d'autres sont en film quasiment continus, comme les couches de passivation (au-dessus).

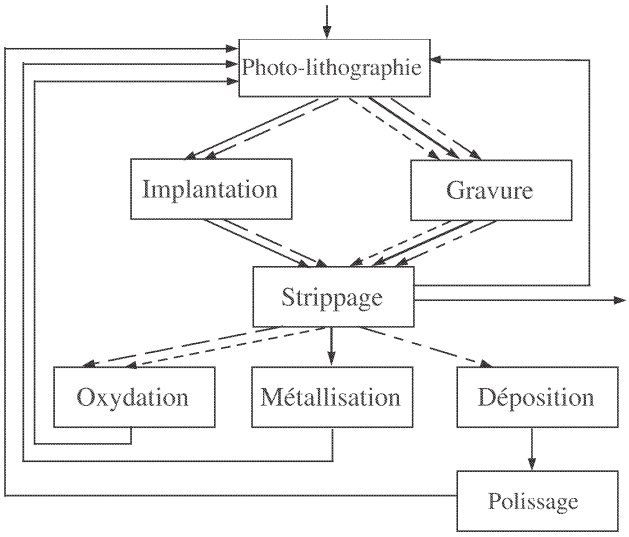


FIG. 2.3 – Séquence des étapes de la fabrication des wafers [120]

La fabrication d'une puce sur un wafer consiste à ajouter, modifier ou enlever du matériau sur des régions spécifiques de la plaquette de silicium. Ces régions peuvent se comporter comme isolant, conducteur ou semi-conducteur afin de former les composants et leurs interconnexions. Ces régions sont fabriquées par une séquence d'opérations, chaque séquence formant une nouvelle couche de matériau sur la plaquette. Le tableau 2.1 présenté pré-

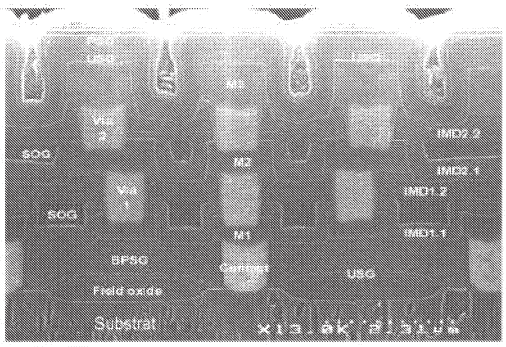


FIG. 2.4 – Image en microscopie électronique d'un composant possédant trois niveaux de métal [102]

cédemment résume les opérations les plus communes de la fabrication des wafers et leur description. Les couches sont fabriquées par étape, et chaque couche nécessite un passage par l'opération de photo-lithographie. Les processus actuels peuvent atteindre les 40 niveaux de masques.

Par le procédé de photo-lithographie, une couche de résine photo-sensible est déposée sur la surface de la plaquette. Ce procédé transpose sur la surface de la plaquette une structure préalablement définie par une masque en utilisant de la lumière ultraviolette, comme illustré dans la figure 2.5. Ainsi, la configuration requise pour chaque couche est obtenue. Les régions non couvertes par la résine pourront être exposées à d'autres traitements. Le procédé de photo-lithographie est usuellement réalisé dans une machine capable d'exécuter les opérations d'application, exposition et développement de la couche photo-sensible. Les machines traitent un wafer à la fois.

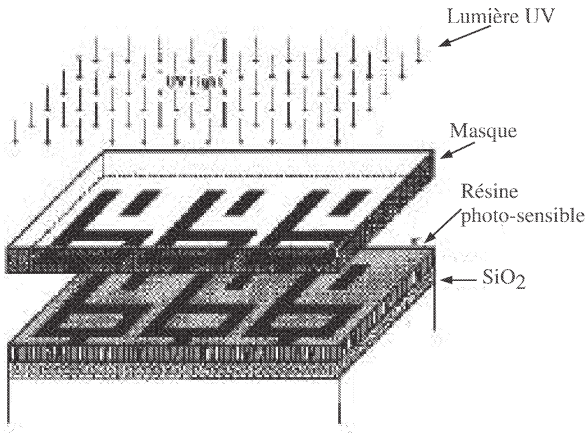


FIG. 2.5 – Concepts de base de la photo-lithographie [120]

Le procédé de gravure sert à enlever du matériau de la surface de la plaquette sur des régions présélectionnées. Le procédé d'implantation ionique est utilisée pour définir des zones

actives. La figure 2.6 montre, respectivement, (a) la surface du wafer couverte de résine, (b) la gravure des régions non couvertes, et (c) l'enlèvement de la résine. A travers l'insertion des dopants, nous pouvons obtenir des régions ionisées soit avec une charge positive soit avec une charge négative. Avant de pouvoir déposer une nouvelle couche de matériau, les résidus des couches précédentes doivent être éliminés. Ceci est réalisé à travers le procédé de strippage.

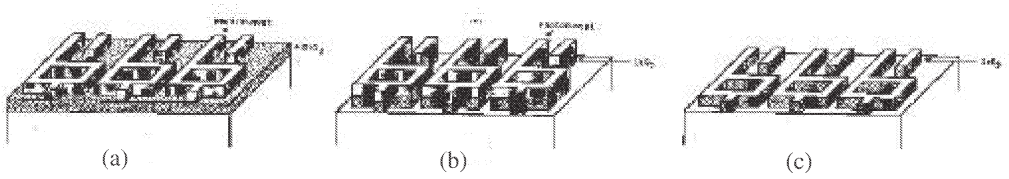


FIG. 2.6 – Les procédés de gravure et strippage [120]

Le procédé d'oxydation thermique est employé pour ajouter une couche de dioxyde de silicium (SiO_2) sur le wafer. Cet oxyde est utilisé comme isolant pour séparer les composants et les connexions, ainsi que comme protection des zones actives. L'oxydation est effectuée dans des fours à très haute température et nécessite une quantité de temps considérable. Par conséquent, les fours sont conçus pour traiter plusieurs wafers à la fois. Ces wafers peuvent appartenir à différents lots.
Le procédé de déposition (*Chemical-vapor deposition*, CVD) est utilisé pour déposer des couches de matériau conducteur ou isolant sur la plaquette. Un procédé appelé *sputtering* (ou *Physical-vapor deposition*, PVD) est également utilisé pour déposer des couches d'un matériau conducteur, ainsi que des isolants. Dans les fabs actuelles, les procédés de CVD et PVD sont exécutés sur des machines de type cluster.
Le polissage mécano-chimique (*Chemical-Mechanical Polishing*, CMP) est utilisé pour aplanir les couches de matériau diélectrique ou métallique. En raison du caractère partiellement mécanique du polissage des plaquettes, des particules polluantes sont produites. Ce procédé doit donc être exécuté dans un endroit isolé des autres ateliers.

Les opérations d'inspection et de défectivité sont fréquemment réalisées pendant tout le processus de fabrication. En raison de la haute sensibilité du produit aux défaillances et du coût très élevé de la fabrication des wafers, il s'avère nécessaire de détecter les défauts très rapidement. Le ratio entre les opérations d'inspection et de défectivité *versus* les opérations de fabrication varie de 50% à 100%.

2.3 L'agencement des machines dans la fab

Le problème d'agencement de machines dans les ateliers (*layout*) a été largement étudié dans la littérature. Pour tout système de fabrication, plusieurs approches ont été proposées [29]. L'approche la plus utilisée est connue sous le nom de Quadratic Assignment Problem (QAP) [13]. Dans le problème QAP standard, l'objectif est de trouver une affectation optimale de N départements (ou machines) à N sites afin de minimiser le coût total

associé à la manipulation et transport des produits. Puisque le QAP est connu comme étant un problème $\mathcal{NP}$-complet [107], les travaux de recherche se sont focalisés sur la proposition d'algorithmes approximatifs ou de méthodes heuristiques efficaces. Une discussion de ces travaux est présentée dans [60].
Toutes ces méthodes peuvent être *a priori* utilisées pour résoudre le problème d'agencement des machines dans une unité de production microélectronique. La complexité du processus de fabrication et les caractéristiques technologiques des équipements ont cependant créé un intérêt particulier de la part des chercheurs. La communauté des semi-conducteurs a ainsi étudié ce problème de façon séparée. Les premiers travaux se sont focalisés exclusivement sur la localisation des machines dans la fab en analysant la viabilité économique de ces configurations par rapport à des indicateurs clés de fabrication (e.g. temps de cycle, débit du système, temps de transport) [125, 36, 38, 32, 17]. Une description détaillée de ces approches est présentée dans [83].
Malgré les avantages potentiels de certaines des configurations proposées, les fabricants de semi-conducteurs sont opposés à essayer de nouveaux *layouts* car la configuration traditionnelle en *bays* (departements ou couloirs) offre plusieurs avantages en termes d'exécution des activités de maintenance et des opérations de production [14]. Ces avantages sont notamment dus à la localisation de machines du même type dans la même *bay*. La figure 2.7 montre deux exemples de configuration par *bay*. L'équipement utilisé dans la fabrication de semi-conducteurs étant très coûteux et très sensible en raison des très petites tolérances requises, les machines sont très souvent à l'arrêt pour maintenance soit suite à une panne ou à un réglage, soit en maintenance préventive. Donc, la localisation des machines du même type dans une même *bay* permet de continuer la production lors de ces périodes de maintenance. Par conséquent, les passages de maintenance peuvent être construits entre les couloirs des ateliers afin de superviser et réparer les machines. Ces couloirs sont ainsi séparés de l'espace dédié à la fabrication et permettent l'accès sans perturber la production. Comme nous le verrons dans la section 2.6, une grande partie de la complexité pour la gestion du transport de produits dans la fab (encours) est fortement due à la configuration des ateliers en *bay*.

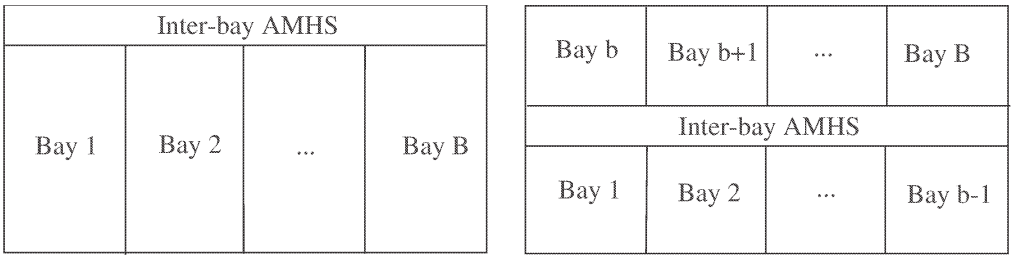


FIG. 2.7 – Configurations possibles d'une fab multi-bay

2.4 Le transport des wafers dans la fab

Bien que traditionnellement dans l'industrie des semi-conduteurs le centre d'intérêt pour l'amélioration des indicateurs de production ait été la gestion des machines, qui sont d'ailleurs très coûteuses et sophistiquées, le transport, la manipulation et le stockage des lots de wafers est aussi crucial pour la performance de la fab. En effet, l'optimisation de la

logistique interne de la fab (i.e. le transport de lots entre les opérations) permet de réduire le temps de cycle et d'augmenter les rendements. Le transport des wafers à l'intérieur de la fab peut être réalisé manuellement par des opérateurs ou par un système automatisé. Dans cette thèse, nous nous intéressons au transport automatisé de wafers.
Depuis quelques années, les fabricants de semi-conducteurs se sont intéressés de près à l'utilisation et l'installation des systèmes automatisés de transport (*Automated Material Handling System*, AMHS). Ils ont dû re-concevoir les activités logistiques de la production et repenser les concepts de stockage et manipulation de lots afin d'améliorer la productivité et la sécurité dans la fab.
D'un point de vue économique, la taille des wafers augmente afin de réduire les coûts associés à la fabrication. Par exemple, une augmentation de 50% du diamètre d'un wafer se traduit par une augmentation de plus de 100% de la quantité de puces par plaquette, lorsque la taille d'une puce est maintenue constante [120]. Ceci nécessite une manipulation plus délicate des lots compte tenu du coût du produit en cours de fabrication. De plus, avec l'augmentation du nombre d'étapes élémentaires de fabrication, un suivi rigoureux des lots dans la fab devient indispensable. Ce suivi peut être assuré grâce à un ordinateur central qui enregistre les mouvements des lots même lorsqu'ils sont transportés ou manipulés par l'AMHS (ce qui n'est pas le cas dans la pratique lorsque les activités de transport et manipulation sont faites manuellement par des opérateurs).
L'installation d'un système automatisé de transport n'est pas seulement primordiale du point de vue économique. Avec l'augmentation de la taille des plaquettes, le poids augmente aussi. Ainsi, une cassette contenant un lot de 25 wafers de la nouvelle génération de 300mm de diamètre pèse environ 12 kg. Donc, afin de diminuer les problèmes ergonomiques dus à la manipulation systématique de boîtes trop lourdes, les nouvelles fabs ont un AMHS pour le transport des lots.
Les systèmes automatiques de transport ont été étudiés dans la littérature. Une description détaillée des besoins technologiques et physiques dans les unités de fabrication de wafers est présentée par Pillai [100]. Cet auteur fournit les lignes directrices qualitatives pour la conception d'un AMHS. Une approche plus quantitative est présentée dans [99] .

Le chapitre 3 de cette thèse présente une discussion plus approfondie de la littérature scientifique sur le transport automatisé de produits dans les unités de production de wafers.

2.5 Mesure de la performance de la fabrication

Dans cette section, nous allons décrire les indicateurs de performance industrielle pertinents pour la fabrication de semi-conducteurs [84]. Parmi ceux-ci, nous avons identifié le temps de cycle de fabrication, le niveau d'encours et le débit du système. Ces indicateurs sont détaillés ci-après.

2.5.1 Temps de cycle

Le temps de cycle de production, appelé aussi temps de séjour ou délai de fabrication, est le temps total nécessaire pour fabriquer un lot de wafers. Il mesure le temps écoulé entre le moment où le lot entre dans la fab et le moment où il sort. Le temps de cycle inclut le temps effectif de fabrication (i.e. le temps total d'exécution des opérations sur les machines), le temps de transfert entre les opérations, le temps d'attente devant les machines et le temps

d'attente pour les transferts. Dans la plupart des unités de fabrication microélectroniques, un lot de wafers peut passer entre 50% et 90% de son temps de cycle en attente d'une machine ou d'un transfert. Le temps de cycle est une mesure de la capacité qu'a la fab de délivrer les produits à temps.
Dans un contexte général, une réduction du temps de cycle équivaut à une augmentation du niveau de satisfaction des clients. Les entreprises avec des temps de cycle courts sont capables de lancer souvent de nouveaux produits, de pénétrer plus rapidement de nouveaux marchés, de réagir plus efficacement aux changements de la demande et de délivrer à temps leurs produits. Elles peuvent aussi offrir à leurs clients des prix plus compétitifs car elles ont rationalisé leurs process avec un niveau d'encours maîtrisé. D'après des études expérimentales, diminuer le temps de cycle de la moitié (et doubler le taux de renouvellement des encours) se traduit par une augmentation de la productivité d'entre 20% et 70% [110]. Avec des temps de cycle réduits, la qualité s'améliore aussi. Lorsqu'on arrive à diminuer les activités de non-valeur ajoutée, il y a moins de probabilité de dommages car les produits sont moins exposés dans les ateliers [39].
Dans l'industrie des semi-conducteurs, la relation $[(\textit{temps de cycle})/(\textit{temps total de process})]$ est appelé "X-factor" [77]. Un objectif essentiel dans la fabrication de semi-conducteurs est d'avoir un X-factor aussi proche de 1 que possible. Cela impose une diminution de la moyenne et de la variance du temps de cycle. Une première façon d'améliorer le temps de cycle consiste à réduire le temps de process (i.e. le temps total de traitement des lots sur les machines). Pour cela, il faudrait améliorer les conditions technologiques des machines, ce qui nécessite un investissement considérable. Donc, une façon plus économique de réduire le temps de cycle consiste à réduire les temps de non-production (attente et transport). Cette réduction est possible grâce à la mise en œuvre de politiques intelligentes pour l'ordonnancement de machines et le pilotage du système de transport.

2.5.2 Niveau d'encours

Le niveau d'encours (*work-in-process*, WIP) est défini comme le nombre de wafers de production se trouvant dans la fab à un instant donné, aussi bien sur une activité de production que sur une activité de non-production (transport et attente).
Comme établie par la loi de Little [73], si un système est stable et stationnaire (i.e. si la probabilité d'être dans un état ne dépend pas du temps), alors la moyenne du WIP est proportionnelle à la moyenne du temps de cycle (CT)

$$WIP = \lambda CT \tag{2.1}$$

Où λ est le taux moyen d'arrivée de lots par unité de temps.

Lorsque λ est fixé, il n'est pas possible de réduire la moyenne du WIP sans réduire la moyenne du temps de cycle et *vice versa*. Il est important de remarquer que cette relation est valable pour toute la fab ou pour une machine donnée. Ainsi donc, un niveau élevé du WIP suite à un temps de fabrication élevé alourdit la charge du système de transport automatique, nécessite d'agrandir les distances entre les équipements pour permettre une meilleure manipulation du WIP et augmente la probabilité de dommages pour le produit [4]. De plus, les encours représentent du capital immobilisé dans les ateliers, ce qui diminue la trésorerie disponible.

2.5.3 Débit du système (throughput rate)

Le débit du système, appelé simplement *"throughput"*, est défini comme le nombre de wafers finis qui sortent de la fab par unité de temps. En général, le *throughput* est calculé pour chaque type de produit (wafer) fabriqué. En plus de donner une idée sur le niveau de saturation ou d'utilisation du système, l'évolution du *throughput* dans le temps permet de savoir si le système est stable, c'est-à-dire, s'il n'y a pas d'accumulation de produits quelque part dans le système (en particulier lorsque le taux d'arrivée de lots est très proche du taux de production) ou si le système évolue selon les prévisions.
Dans une ligne de fabrication idéale, sans aucune variabilité de temps d'arrivée ou de process, les produits n'ont pas de temps d'attente lorsque le niveau d'utilisation est inférieur à 100%. Avec un niveau d'encours bas, le temps de cycle est égal au temps de process total théorique plus le temps de transport. Tant que le débit est inférieur à la capacité de la ressource goulot, le temps de cycle reste à cette valeur minimale, sans aucun temps d'attente. A partir du moment où le *throughput* arrive à sa valeur maximale, les temps d'attente commencent à apparaître. A partir de ce point, une augmentation du niveau des encours se traduit par une augmentation du temps de cycle causée par l'attente devant les machines [120]. Cette situation théorique est illustrée dans la figure 2.8.

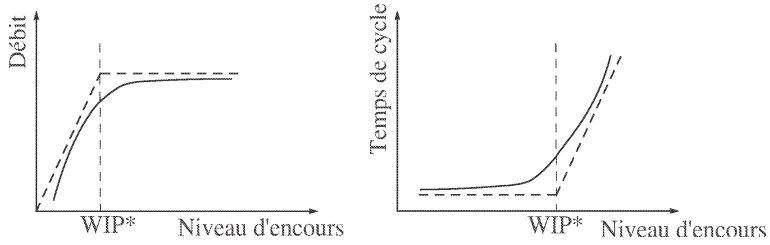


FIG. 2.8 – Encours vs. temps de cycle et throughput

Le WIP optimal (WIP^*) est le point où le *throughput* arrive à sa valeur maximale, tandis que le temps de cycle est encore à sa valeur minimale. WIP^* est ainsi appelé point critique et est calculé comme :

$$WIP^* = \delta_{max} CT_{min} \tag{2.2}$$

Où δ_{max} est le débit (*throughput*) maximum.

En conséquence, le *throughput* théorique et le temps de cycle dans le cas de variabilité zéro sont :

$$\delta_0(WIP) = \begin{cases} \frac{WIP}{CT_{min}} & \text{si } WIP < WIP^* \\ \delta_{max} & \text{si } WIP \geq WIP^* \end{cases} \tag{2.3}$$

$$CT_0(WIP) = \begin{cases} CT_{min} & \text{si } WIP < WIP^* \\ \frac{WIP}{\delta_{max}} & \text{si } WIP \geq WIP^* \end{cases} \tag{2.4}$$

La ligne de flux idéale fournit les asymptotes pour les situations non-idéales où la variabilité est présente. En raison de la variabilité affectant les temps inter-arrivées et les temps opératoires (ou temps de process), les courbes du temps de cycle et du *throughput* ressemblent à celles de la figure 2.8. Le temps de cycle minimal est ainsi égal à la somme du temps de process plus le temps de transport. Lorsque les niveaux d'encours augmentent, même si le taux d'utilisation est inférieur à 100%, les temps d'attente commencent à apparaître, comme indiqué par les équations. Ces temps d'attente sont le résultat de la variabilité dans le système. Dans la situation non-idéale, il n'y pas de point critique pour le WIP, mais un intervalle critique. Dans cet intervalle, on a le choix entre un niveau maximal pour le *throughput* avec un certain temps de cycle ou un temps de cycle plus court mais avec un *throughput* inférieur à l'optimum.
Glassey et Resende [33] ont observé qu'il existe une relation particulière entre l'augmentation du *throughput* et la capacité de production d'une fab : le temps de cycle tend vers l'infini lorsque le *throughput* s'approche de la capacité maximale de la fab (voir figure 2.9). Cette courbe est appelée la courbe opératoire (*operating curve*) [5], qui est très souvent utilisée par les fabricants comme un indicateur clés de la productivité de la fab.

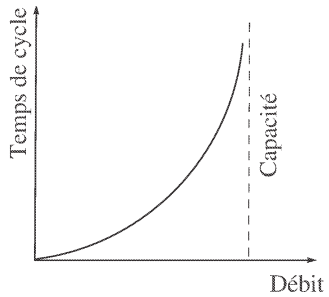


FIG. 2.9 – Relation entre temps de cycle et throughput

2.6 Difficultés pour la gestion des opérations

La gestion des opérations fait référence à la gestion du flux des produits et aux opérations de transformation des matières premières en produits finis. Dans une unité de fabrication de semi-conducteurs, il y a plusieurs facteurs qui en font un environnement particulièrement difficile. Le lecteur intéressé peut se référer aux travaux de Hogg *et al.* [40] ou Uzsoy *et al.* [118, 119] pour une discussion détaillée sur ce sujet. Cette section présente les principaux aspects liés à cette complexité.

Comme expliqué précédemment, le processus de fabrication de wafers est réalisé couche par couche. La réalisation de chaque couche nécessite l'exécution de plusieurs étapes individuelles (photolithographie, gravure, etc.). De plus, beaucoup de ces étapes sont répétées plusieurs fois. Les machines utilisées pour les exécuter sont très coûteuses et ne sont donc pas multipliées dans les ateliers. Ainsi, les lots doivent passer plusieurs fois sur une même séquence de machines. Le produit fini peut donc être vu comme étant un "sandwich multicouches". Les systèmes de fabrication dans lesquels les lots doivent visiter plusieurs fois les mêmes machines à des étapes différentes de la gamme opératoire sont appelés lignes

re-entrantes [57]. La principale conséquence du caractère re-entrant est que des wafers à différents stades du cycle de fabrication sont en compétition pour la même machine. De cette façon, les plaquettes passent une grande partie de leur temps soit en attente d'une machine, soit dans une activité de transport entre deux opérations, au lieu d'être effectivement traitées sur une machine.
Cette caractéristique principale engendre des problèmes d'ordonnancement et de gestion des opérations particulièrement difficile dans la pratique et théoriquement intraitables (i.e. problème d'instabilité du système re-entrant) [57, 58].

En plus de la nature re-entrante du processus de fabrication de wafers, d'autres facteurs rendent la gestion des opérations particulièrement difficile [118, 119, 47, 25] :

- **Rendements aléatoires :** Les rendements du processus de fabrication sont incertains et peuvent varier du fait de problèmes sur les équipements de production ou les conditions environnementales de la salle blanche. Lors de l'étape de test, il y a aussi des conditions non-optimales qui peuvent dégrader le produit. Les prévisions des rendements de la fabrication pour les produits bien établis ne sont pas difficiles puisqu'on peut utiliser des données historiques. Cependant, la constante introduction de nouveaux produits et de nouvelles technologies en fait un problème très complexe.
- **Caractéristiques diverses des équipements :** Les caractéristiques des équipements utilisés pour la fabrication de wafers sont très diverses : machines avec temps de préparation (*set-up times*) très variables, machines à traitement par batch, machines séquentielles, etc.
- **Gestion de la maintenance :** Comme indiqué précédemment, la fabrication de wafers est un environnement *high tech* où les machines utilisées sont extrêmement sophistiquées. Donc, même si un calendrier de maintenance préventive est bien établi, elles sont très souvent à l'arrêt suite à une panne inattendue ou en calibrage suite à un produit hors spécifications.
- **Production et développement partagé dans les fabs :** En raison du développement constant de nouveaux produits et procédés, les mêmes équipements dans la fab sont utilisés à la fois pour le traitement de wafers dédiés à la production (i.e. livré aux clients), que pour les wafers dédiés à la recherche et développement et d'ingénierie pour le test et la qualification des machines. Cette confusion augmente à cause du conflit entre les objectifs de la production et des ingénieurs de développement.
- **Disponibilité et maintenance des données de la production :** Du fait de la quantité de données à traiter dans les fabs, l'obtention et la gestion de l'information sont des tâches très compliquées et ardues [18].

2.7 Conclusion

Dans ce chapitre, nous avons présenté les principaux aspects de la fabrication de wafers. Une telle discussion est importante afin de mieux comprendre les caractéristiques de ce secteur industriel et la difficulté de gestion pratique de ce type de fabrication. Le chapitre suivant se focalise sur le rôle du système automatisé de transport de produits dans la fabrication de wafers. Nous y ferons une analyse critique de la littérature afin d'identifier l'objectif de cette thèse.

Chapitre 3

Le transport automatisé de wafers

Ce chapitre est dédié à l'analyse de la littérature sur la conception et la gestion des systèmes automatisés de transport dans les unités de production de semi-conducteurs. D'abord, nous présentons quelques notions générales sur les systèmes automatisés pour le transport de produits (AMHS). Ensuite, nous analysons la littérature scientifique sur le problème de conception du réseau et sur le pilotage de tels systèmes et identifions ainsi ses manques. Le positionnement de nos travaux dans la littérature et notre démarche de recherche sont présentés à la fin du chapitre.

3.1 Notions préliminaires

Avec l'arrivée de wafers de 300mm de diamètre, les unités modernes de fabrication de wafers nécessitent un niveau d'automatisation beaucoup plus élevé qu'auparavant. La littérature scientifique comporte depuis les années 1980 de nombreux travaux traitant des problèmes de conception et de gestion dans les systèmes automatisés de distribution et de stockage et les systèmes flexibles avec chariots autoguidés (AGV) [78, 28, 55]. Un état de l'art récent est présenté dans [101].

Pour la conception et la gestion d'un système automatisé de transport, plusieurs problèmes se posent. Les types de décisions qui doivent être prises peuvent être classées en trois catégories classiques : stratégiques, tactiques et opérationnelles. Au niveau stratégique, la première étape consiste à définir les contraintes sur le long terme et les indicateurs de performance les plus pertinents, comme par exemple l'estimation des flux de produits et le type de système à installer (i.e. technologie : convoyeur, véhicules autoguidés, convoyeur aérien). Au niveau tactique, il est nécessaire de définir la configuration du réseau (e.g. localisation des points de chargement et déchargement) et d'estimer le nombre de véhicules nécessaires dans le système. Le niveau opérationnel fait référence aux décisions opérationnelles quotidiennes comme le dispatching et le routage de véhicules, c'est-à-dire, quel véhicule est affecté à quelle requête (ou *vice versa*) et quand cette requête sera servie.

Il y a une grande interaction entre les niveaux stratégique, tactique et opérationnel [121]. Par exemple, le nombre de véhicules dépend de la stratégie de gestion implémentée (règles de dispatching/positionnement), et le pilotage des véhicules dépend des restrictions du système, des indicateurs de performance et du *layout*. De plus, contrairement à ce qui est traditionnellement accepté dans la littérature, les décisions au niveau opérationnel auront un grand impact sur les décisions des niveaux supérieurs. Par exemple, afin de justifier une

réduction du nombre de véhicules, il s'avère nécessaire d'estimer l'impact sur les indicateurs de performance et les temps de réponse. S'il y a trop de véhicules dans le système, alors le capital investi est important et il y a une plus forte probabilité de créer des problèmes de congestion dans le réseau, rendant le temps de transport très important. Au contraire, s'il n'y a pas assez de véhicules par rapport au niveau de fonctionnement des machines de production, les produits vont saturer les stocks de sortie des machines, les temps d'attente pour un véhicule seront très longs et les délais de livraison sur les machines ne seront pas respectés.
Or, une approche de résolution intégrant les trois niveaux semble très difficile à mettre en place. C'est pourquoi une approche séquentielle est souvent utilisée, où les décisions stratégiques sont d'abord prises en considérant des données agrégées provenant du niveau tactique. Ensuite, les décisions tactiques et opérationnelles sont affinées à partir d'informations plus précises.

Enfin, quel que soit le principe de transport utilisé, l'AMHS doit permettre une grande flexibilité et satisfaire le débit demandé du système. Ainsi, le choix du mode de transport est fait à partir des objectifs visés par l'entreprise pour les long et moyen termes. Dans les fabs de wafers, la contrainte vient essentiellement de l'organisation des ateliers en *bays* (voir section 2.3). Le choix de la structure du système de transport interne est extrêmement important et influe sur la performance du système entier. Par la suite nous présentons une analyse de la littérature.

3.2 Conception de l'AMHS pour les fabs

L'objectif de cette thèse étant l'étude de stratégies de pilotage du système de transport automatisé, rappelons que la configuration du réseau de transport et l'agencement des machines ont un grand impact sur ses performances [100]. Cette section présente une revue de la littérature sur la conception du réseau et la sélection du type d'AMHS (technologie) pour la fabrication de wafers. Un état de l'art plus détaillé est présenté dans [83]. La détermination du nombre de véhicules nécessaires pour le transfert des lots est un autre problème qui doit également être traité lors de l'étape de conception de l'AMHS.
Les opérations de transport de wafers dans une fab peuvent fondamentalement être classées en opération de type inter-bay (entre les *bays* ou départements de fabrication) et intra-bay (à l'intérieur d'une *bay*). Dans la littérature, la plupart des travaux se focalisent sur la configuration de systèmes de transport inter-bay. Chronologiquement, l'intérêt pour les systèmes intra-bay n'est devenu pertinent dans la pratique que depuis très peu. Cette classification en "systèmes inter-bay" et "systèmes intra-bay" a aussi été reprise pour approcher le problème de pilotage des véhicules. Il s'avère donc important de comprendre l'évolution chronologique des AMHS dans les fabs : système inter-bay, système séparé inter-bay/intra-bay, système inter-bay/intra-bay interconnecté mais non-unifié et aujourd'hui système complètement unifié.

3.2.1 Système de transport inter-bay

L'objectif d'un système automatisé de transport inter-bay est de transporter le produit encours de fabrication entre deux départements de production (*bay*) différents sans l'intervention des opérateurs. Typiquement, le système nécessite impérativement des étagères

pour le stockage temporaire des lots localisées à l'extérieur de chaque *bay*, dont la taille est calculée afin de pouvoir stocker le produit en cours (WIP) de la *bay* correspondante [100]. Dans la littérature, plusieurs auteurs se sont intéressés au problème.
Peters et Yang [98] ont étudié les configurations de type *spine* (en forme de colonne vertébrale) et en périmètre (voir figure 3.1). Dans une configuration de type *spine*, une seule boucle uni-directionnelle est fixée dans la fab, usuellement dans le couloir central, de telle façon que toutes les *bays* puissent être desservies par les véhicules. Des croisements (*turntables*) doivent être installés le long de la boucle afin de permettre aux véhicules de changer de sens de direction et délivrer les cassettes. Dans la configuration en périmètre, le circuit est composé de deux boucles séparées qui contournent la fab, chacune ayant son propre sens de circulation pour les véhicules [15]. Dans la méthode proposée par Peters et Yang, le *layout* de la fab étant préalablement défini, le problème de configuration de l'AMHS est résolu à partir d'une solution initiale réalisable et la valeur de la fonction objectif est améliorée en évaluant la réduction du coût associé à un échange de deux départements dans la séquence du flux. Un inconvénient de cette méthode est sa sensibilité à la solution initiale. Afin de résoudre ce problème, Yang *et al.* [128] ont proposé un algorithme hybride recherche tabou-recuit simulé qui est moins sensible à la séquence initiale des flux et aussi plus efficace en temps de calcul. En outre, pour la configuration *spine*, une approche intégrée basée sur la résolution du Quadratic Set-Covering Problem (QSCP) peut être utilisée [127]. Cette procédure exploite la configuration particulière des fabs de semi-conducteurs pour déterminer la configuration optimale de l'AMHS. Ce modèle considère que les coûts de transport sont directement proportionnels à la distance de déplacement d'un véhicule. L'objectif est donc de trouver la configuration du réseau (i.e. localisation des étagères) qui minimise le coût total de transport. D'autres configurations telles que les configurations circulaire et *spine* bi-directionnel ont été également étudiées [113, 114]. Les auteurs utilisent des techniques de la Recherche Opérationnelle (e.g. programmation dynamique, théorie de graphes) couplées avec des modèles de simulation à événements discrets pour résoudre le problème. Ces modèles cependant sont limités quant à leur applicabilité générale car ils considèrent des fabs relativement petites (i.e. peu de points de chargement/déchargement). De plus, les pannes de l'AMHS ne sont pas prises en compte. Ce dernier point est essentiel car l'AMHS doit être conçu de telle façon que si un tronçon ou une partie du système tombe en panne, le temps de cycle de fabrication n'augmente pas à cause du blocage du réseau.

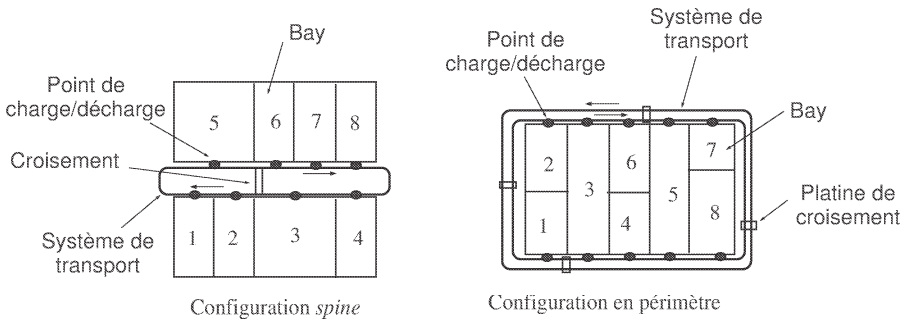


FIG. 3.1 – Configurations *spine* et en périmètre

3.2.2 Système de transport intra-bay

Les transferts intra-bay d'un système automatisé de transport font référence aux mouvements entre les zones de stockage intermédiaires (étagères ou stockers) où attendent les lots et les équipements de fabrication (machines) [100]. Ceci implique qu'il y a des points de chargement/déchargement pour chaque équipement automatisé où les véhicules pourront prendre/déposer les cassettes avec les lots à transporter. Malgré son importance dans la pratique industrielle actuelle, peu de travaux aujourd'hui traitent dans la littérature de ce problème, et les travaux réalisés considèrent des configurations particulières de fabs spécifiques avec des systèmes de transport non-unifiés (i.e. réseaux de transport où les boucles inter-bay et intra-bay sont physiquement séparés) [37, 75, 76, 69, 71]. Le transfert d'un lot entre la boucle inter-bay et la boucle intra-bay est donc réalisé à l'aide d'étagères qui servent de connexion entre les deux configurations.
Les résultats de ces travaux sont donc difficilement applicables pour les AMHS des nouvelles fabs de 300mm où le système est complètement unifié, c'est-à-dire que les boucles sont parfaitement interconnectées autorisant les véhicules à réaliser des mouvements directement d'une machine vers une autre sans les obliger à déposer la cassette sur une étagère de stockage intermédiaire.

3.2.3 Taille de la flotte de véhicules

Une étape importante de la conception d'un système automatisé de transport est la détermination du nombre de véhicules. La taille de la flotte de chariots est un facteur déterminant pour la performance du système [8, 31]. Comme expliqué précédemment, s'il est insuffisant par rapport au niveau de fonctionnement des machines de production, les produits vont saturer les stocks en sortie des machines, attendant qu'un véhicule vienne les chercher. S'il est trop élevé, les chariots vont congestionner le réseau, rendant les temps de transport très importants. Il existe dans la littérature sur les systèmes flexibles de production (FMS) avec véhicules autoguidés (AGV), un grand nombre de travaux traitant de la problématique de détermination de la taille de la flotte de véhicules. Ces travaux peuvent être classés en trois catégories : les méthodes analytiques, les méthodes basées sur la simulation, et les méthodes hybrides (analytiques avec simulation). Une revue de la littérature est présentée dans [65].
Ces approches peuvent *a priori* être utilisées pour tout réseau de transport automatisé avec des mouvement horizontaux. Néanmoins, lorsqu'on veut appliquer certaines de ces méthodes pour une fab de semi-conducteurs, au moins deux inconvénients se présentent. D'une part, en raison de la taille du système et de la complexité de la fabrication, le temps de calcul devient très important [85]. D'autre part, pour une configuration du réseau et un volume de production donnés, le besoin en nombre de véhicules dépend fortement de la (des) politique(s) de gestion opérationnelle mise(s) en place [16, 87].
Dans la littérature des semi-conducteurs, peu de travaux traitent du problème de calcul du nombre de véhicules. A l'exception des travaux de Rooney *et al.* [105] et Sturm *et al.* [111], les autres travaux publiés considèrent une taille de flotte préalablement définie dans la simulation. Rooney *et al.* [105] ont explicitement calculé le nombre de véhicules nécessaires dans un système de transport inter-bay à partir d'un modèle de fab statique déterministe. Une fois ce calcul réalisé, les caractéristiques dynamiques du système (e.g. temps de chargement/déchargement non-déterministes) ont été pris en compte dans un

modèle de simulation, dont une analyse de sensibilité permettait d'étudier l'impact de ce facteur sur les performances de l'AMHS. Le travail de Sturm *et al.* [111] propose une méthodologie pour calculer le nombre de véhicules lors de la planification de l'AMHS basé sur l'analyse du niveau de congestion. Encore une fois, ces travaux considèrent des systèmes non-unifiés. Le calcul est donc réalisé indépendamment pour chacune des boucles du réseau de transport lesquelles ne sont pas interconnectées.

3.3 Le pilotage du système de transport

Un système automatisé de transport et de manutention est composé de deux sous-systèmes interactifs : matériel et opérationnel (*hardware/software*). Le coté matériel est lié aux composants physiques du système tels que les véhicules, le circuit de guidage, etc. La partie opérationnelle englobe les approches et algorithmes pour contrôler et coordonner de manière automatique les ressources du niveau matériel de l'AMHS global, de sorte que ce dernier fonctionne de manière harmonieuse et efficace.
Dans la littérature des semi-conducteurs, la partie matérielle est bien développée et maîtrisée depuis quelques années (voir section 3.2), alors que la partie opérationnelle ne s'est développée que tout dernièrement (voir section 3.4). Ce dernier cas est aussi valable pour les systèmes automatisés installés dans d'autres systèmes de fabrication ou de manutention, où les chercheurs sont confrontés à des problèmes critiques, comme par exemple le transport par conteneurs dans les ports, avec un grand nombre de transporteurs (de l'ordre d'une centaine) et de tâches à réaliser [101]. Plusieurs problèmes tels que la congestion ou les interblocages ont lieu pendant le fonctionnement d'un système avec véhicules automatisés. Ils doivent être pris en compte dans l'étape de la commande en temps réel afin d'assurer le meilleur fonctionnement du système.
La littérature a toujours considéré que la partie opérationnelle est composée de deux problèmes importants : le problème de dispatching et le problème de routage. Nous verrons cependant plus loin dans cette thèse (chapitres 5 et 6) que d'autres problèmes tels que la recherche de tâches de transport ou encore le positionnement de véhicules se posent également au niveau opérationnel. A notre connaissance, ces deux derniers problèmes n'ont pas été étudiés auparavant. La figure 3.2 présente les relations entre les différents aspects du pilotage opérationnel du système de transport automatisé.

3.3.1 Le problème de dispatching

Le dispatching est en fait un ordonnancement des véhicules. Lorsqu'une opération d'un lot se termine sur une machine, celle-ci émet une requête de transport, c'est-à-dire une demande pour transporter le lot vers sa prochaine destination dans la gamme. Une requête émise est dite en attente tant qu'un véhicule libre ne lui a pas été affecté. Une politique de dispatching consiste alors à choisir une requête parmi celles en attente, et aussi à choisir un véhicule capable de réaliser le déplacement désiré. Egbelu et Tanchoco [28] ont mis en évidence les principales règles de dispatching à travers deux situations particulières. La première situation représente le cas où un seul poste de travail émet des requêtes. Dans ce cas, le dispatching consiste à choisir un véhicule parmi une flotte de chariots libres (*vehicle selection problem*). Cela conduit à des politiques dites à l'initiative de la machine. Ci-dessous sont données les politiques décrites dans [28] :

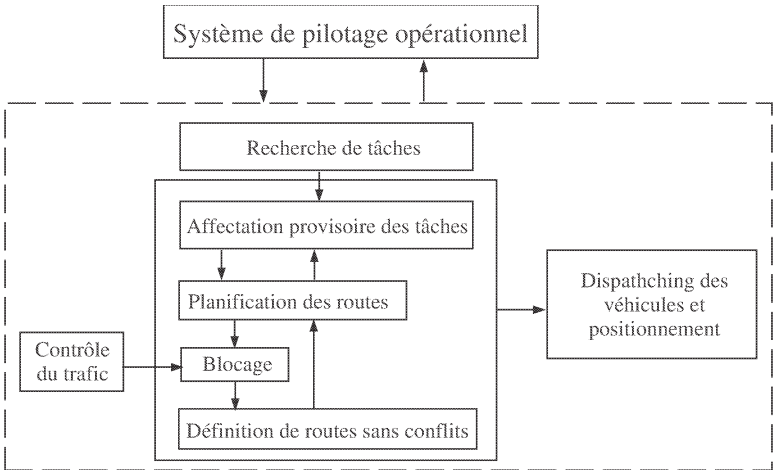


FIG. 3.2 – Pilotage opérationnel d'un système automatisé de transport

- RV (*Random Vehicle*) : Un chariot libre est choisi aléatoirement et est affecté à la station ayant émis la requête de transport. Cette méthode étant neutre dans le choix des véhicules, elle est utilisée comme élément de comparaison.
- NV (*Nearest Vehicle*) : Dès l'émission d'une requête, les distances séparant les véhicules libres et la station demandeuse sont évaluées, et le véhicule choisi est celui dont la distance est la plus petite. On utilise donc la proximité en termes de distance. Si on prend en compte la vitesse du chariot, on peut appliquer alors la proximité temporelle, et le véhicule choisi sera celui qui requiert le moins de temps pour arriver à la station demandeuse. Cette règle permet donc de minimiser les distances parcourues à vide, car elle sélectionne en priorité le couple <véhicule, produit> dont la phase de récuperation est la plus courte.
- FV (*Farthest Vehicle*) : Le véhicule libre choisi sera celui qui est le plus éloigné de la demande (éloignement en termes de distance ou de temps). D'un point de vue opérationnel, cette règle ne présente pas vraiment d'intérêt, mais du point de vue de la conception, elle donne une référence pour l'amélioration de l'efficacité du système (en quantifiant les effets des voyages inutiles des véhicules).
- LIV (*Longest Idle Vehicle*) : Le chariot qui est resté inactif le plus longtemps sera le véhicule affecté. Cette règle assure un équilibre de la charge de travail pour tous les véhicules dans le système.
- LUV (*Least Utilized Vehicle*) : Le véhicule sélectionné sera celui qui aura le moins servi à l'instant de la prise de décision, c'est-à-dire celui dont le taux d'utilisation est le plus faible.

La seconde situation représente le cas où il y a un seul véhicule libre et plusieurs stations qui ont émis des requêtes. Ici, le choix consiste à sélectionner une machine ou une charge parmi plusieurs en attente d'un chariot (*load selection problem*). On a alors des politiques de dispatching dites à l'initiative du véhicule. Dans cette catégorie, nous avons les règles suivantes [28] :

- RL (*Random Load*) : Dès qu'un chariot devient libre, on choisit aléatoirement l'entité qui sera servie.

- STT/D (*Shortest Travel Time/Distance*) : La station desservie sera celle qui est la plus proche du véhicule en termes de temps ou de distance de trajet.
- LTT/D (*Longest Travel Time/Distance*) : C'est la règle inverse de la précédente. D'un point de vue opérationnel, cette règle n'a aucun intérêt. Cependant, du point de vue de la conception, cette règle permet de réaliser des expériences comparatives sur le système.
- FCFS (*First Come First Serve*) ou FIFO (*First In First Out*) : Le chariot est affecté à la requête qui attend depuis le plus longtemps.
- ULSAT (*Unit Load Shop Arrival Time*) : Le véhicule dessert la station dont la pièce devant être transportée est entrée en fabrication depuis le plus longtemps.

Ces règles ne couvrent pas tous les cas pouvant être rencontrés, par exemple le cas de plusieurs requêtes et plusieurs chariots. Une technique permettant le dispatching de plusieurs requêtes et plusieurs chariots est donnée dans [1].

3.3.2 Le problème de routage

Une fois qu'un véhicule est affecté à une mission de transport, l'objectif du routage est de trouver un trajet optimal (e.g. le plus court en termes de distance ou de temps) et réalisable, qui le mène de sa position actuelle à sa destination finale. Une décision de routage comprend trois aspects. D'abord, elle doit chercher l'existence d'un chemin qui mènera le chariot de son origine vers sa destination. Ensuite, le chemin doit être réalisable, c'est-à-dire libre de conflits et de blocages et sans congestion. Enfin, ce chemin doit être optimal (ou pseudo-optimal) selon une fonction objectif [101].
Outre les systèmes de fabrication avec transport automatisé au moyen de véhicules, le problème de routage se pose dans plusieurs domaines tels que l'informatique (routage de paquet d'informations à travers un réseau) et le domaine de la logistique et des transports routiers et ferroviaires [101].

3.3.3 Le positionnement des véhicules

Le problème de positionnement des véhicules est un fait un problème de localistion. Dans la littérature, les modèles proposés se sont principalement intéressés par la localisation d'installations (e.g. usines, entrepôts) qui doivent servir des points de demandes (e.g. clients) [22]. Ces modèles peuvent bien sur être adaptés pour résoudre d'autres problèmes similares. En particuler, une approche du problème de positionnement de véhicules concerne la problématique de localisation d'ambulances et de voitures de pompiers. Dans ce cas, il s'agit d'une problématique retrouvée dans les services publics, et par conséquent, l'objectif est de maximiser la couverture tout en garantisant un temps de service (arrivée sur le site de demande dans un intervalle de temps prédéfini) [12]. Les approches utilisées pour ce cas se basent sur les modèles classiques du *Maximal Covering Location Problem.*

Dans la littérature des systèmes automatisés de production, le problème de positionnement de véhicules (*vehicle positioning problem*) a traditionnellement consisté à trouver un emplacement des chariots lorsqu'ils se libèrent et qu'il n'y a pas de requête en attente. Au lieu de diriger les véhicules vers un dépôt, l'objectif est de trouver des points sur le réseau où les stationner afin qu'ils restent le plus près possible des machines pour pouvoir satisfaire une demande de transport lorsque celle-ci arrivera. Plusieurs approches ont été proposées dans la littérature, principalement focalisée sur une configuration du circuit avec une boucle

simple uni-directionnelle ou bi-directionnelle où les points de chargement/déchargement sont localisés le long du circuit (e.g. [27, 52, 30]). Dans ces travaux, l'objectif est de minimiser le temps d'attente entre le lancement de la requête et l'arrivée d'un chariot.

Dans cette thèse, nous considérons une autre approche du problème de positionnement de véhicules dans la fab différente de celles précédemment décrites. L'intérêt de notre problématique est dû au caractère unifié de l'AMHS dans la fab, les véhicules étant autorisés à effectuer des livraisons directement d'une machine vers une autre, comme expliqué précédemment. Notre objectif est de distribuer les véhicules dans le réseau de telle façon que le système soit couvert et les requêtes satisfaites dans les meilleurs délais. Il s'agit donc de toujours avoir des véhicules le plus proche possible des endroits où il y aura une demande de transport. D'un point de vue formel, nous verrons au chapitre 6 que notre objectif ressemble à celui du problème du P-center (nous voulons minimiser le temps maximal pour couvrir une demande). Les différences entre notre approche du problème de positionnement et celles précédemment décrites et le P-center standard concernent l'objectif du modèle et les contraintes prises en compte. Dans le problème P-center, l'objectif est de trouver la localisation de P installations qui minimise la distance (ou le temps) maximale de couverture. Les P installations sont en fait les ressources qu'il faut distribuer pour satisfaire les demandes des clients. Dans notre cas, nous voulons créer P zones (i.e. diviser le réseau de transport en P zones). Pour nous, la demande des points du réseau (les clients dans le P-center) est satisfaite par une flotte de V véhicules qui doivent être distribués de façon à respecter la capacité globale du système (i.e. la capacité totale de transport). Aussi, la prise en compte des contraintes de capacité nous rapproche du *Capacitated Facility Location Problem* (CFLP), mais nous n'avons pas de coûts liés à l'ouverture d'une installation (i.e. choix d'un centre de zone). De plus, dans notre cas nous avons toujours la certitude d'avoir des capacité discrètes car nos ressources sont en fait les véhicules disponibles qu'il faut repartir dans les différentes zones créées. Cela va nous permettre d'ajouter des critères d'optimisation dans la fonction objectif afin d'équilibrer les demandes et d'équilibrer les ressources (véhicules) dans tout le réseau.

Bien que cette approche du problème de positionnement de véhicules soit inspirée d'un cas industriel provenant de l'industrie des semi-conducteurs, il est tout à fait possible de la retrouver dans d'autres systèmes nécessitant une logistique automatisée (e.g. fabrication automatisée avec véhicules auto-guidés ou encore le chargement, déchargement et transport dans les ports maritimes). Comme nous le verrons au chapitre 5, notre problème de positionnement est différent de ceux existants dans la littérature en raison de la taille du réseau transport (i.e. nombre de points dans le réseau, nombre de véhicules).

3.3.4 La recherche de tâches

Le problème de recherche de tâches (*task-search problem*) consiste à définir l'espace de couverture dans lequel les véhicules, une fois disponibles, doivent rechercher les lots demandant un service de transport [87]. Dans le passé, la littérature spécialisée ne s'est pas intéressée à ce problème parce que les réseaux de transport dans les systèmes de production automatisés étaient assez simples (e.g. une seule boucle) et ne comprenaient qu'un nombre réduit de véhicules. De plus, la littérature des semi-conducteurs n'a pas non plus étudié cette question car les AMHS installés étaient non-unifiés : un chariot ne pouvait pas sortir de

TAB. 3.1 – Stratégies de pilotage des véhicules

Référence	**Configuration de l'AMHS**	**Stratégie**	**Fonction(s) objectif**
Rooney *et al.* [105]	Inter-bay	FCFS, NV	Déplacements par unité de temps
DeJong et Wu [23]	Inter-bay, Intra-bay (non-unifié)	FCFS, NV	Temps de livraison, Impact des *hot lots*
Liao et Wang [68]	Inter-bay, Intra-bay (non-unifié)	FCFS, neural networks	Temps de livraison, Impact des *hot lots*
Jimenez *et al.* [46]	Inter-bay, Inter-floor	FCFS, NV	Temps de livraison
Lin *et al.* [70]	Inter-bay	FEFS, FCFS, NV	Temps de transport et d'attente, Nombre de livraison par unité de temps, Utilisation des véhicules
Jang *et al.* [44]	Inter-bay	Règle anticipatrice	Temps de cycle, Utilisation des véhicules

sa boucle et les lots devaient passer par un stockage intermédiaire avant d'être délivrés [91].

Dans les unités modernes de production de wafers, l'AMHS est un système unifié dans lequel les véhicules sont autorisés à sortir de leurs boucles afin de délivrer les lots, éventuellement directement d'une machine à une autre sans forcément devoir passer par un stockage intermédiaire. Ainsi, le besoin à la fois de simplifier les tâches d'affectation de véhicules aux demandes et de réduire les coûts associés à la gestion des chariots dans le système fait appel à des stratégies plus sophistiquées et intelligentes de gestion de l'AMHS.

3.4 Travaux sur la gestion des véhicules dans les fabs

Les travaux qui existent dans la littérature sur le transport dans l'industrie des semi-conducteurs se sont uniquement intéressés au problème de dispatching des véhicules. Les règles de gestion du système de transport implémentées sont en fait de règles de "bon sens" sans une véritable comparaison expérimentale de leur performance et efficacité. Un résumé de ces travaux est présenté dans le tableau 3.1. Ils s'intéressent à l'étude des conditions pour lesquelles les règles testées fonctionnent bien à l'aide de modèles *ad hoc* de simulation à événements discrets. En fait, la simulation est préférée aux modèles analytiques, car elle permet de prendre en compte l'évolution dynamique de la fab.

La règle la plus utilisée est FCFS (*First Come First Serve*) avec éventuellement une prise en compte des lots prioritaires (ou *hot lots*). Cette règle affecte les chariots aux demandes dans l'ordre d'arrivée dans la file d'attente. FCFS a été utilisée dans les travaux de plusieurs auteurs [105, 23, 46, 68] qui analysent les performances de l'AMHS indépendemment de son impact sur celles du système de fabrication.

Considérant l'AMHS comme un système isolé de la fabrication, Lin *et al.* [70] ont montré que la combinaison des règles NV (*Nearest Vehicle*) et FEFS (*First Encountered First Served*) peut améliorer la performance du système de transport en termes de temps d'attente d'un véhicule et de taux d'utilisation de ces derniers. Pour leur part, Jang *et al.* [44] ont testé une règle anticipatrice sur une configuration inter-bay très simple. Les résultats de leurs simulations montrent une amélioration des indicateurs de performance par rapport aux règles réactives type FCFS. Les auteurs suggèrent que leur règle peut être très efficace dans les systèmes où le nombre de véhicules n'est pas un facteur critique. Cependant, dans les fabs avec un système AMHS unifié, le nombre de véhicules devient un facteur critique car ceux-ci sont très chers, et l'objectif est donc de les gérer de façon efficace et ainsi réduire la taille de la flotte. De plus, leur étude considère uniquement un système avec boucle simple pour le transport inter-bay des wafers. Par conséquent, l'efficacité de cette règle dans un AMHS unifié intégrant déplacements inter-bay et intra-bay n'est pas évidente.

Comme nous le verrons dans la suite de notre travail, la mise en place d'une règle telle que FCFS n'est pas forcément le meilleur choix étant donné le fait que notre AMHS est un système unifié. Ceci implique également qu'une gestion intelligente des véhicules s'impose si nous souhaitons diminuer les coûts d'investissement associés tout en gardant la même qualité de service de transport.

3.5 Synthèse et démarche

Dans ce chapitre, nous avons vu que l'automatisation du transport et de la manutention de produits dans les systèmes de fabrication n'est pas un nouveau concept, les premiers travaux datant des années 1980. Dans l'industrie des semi-conducteurs, alors que l'automatisation des opérations de fabrication a été une priorité depuis toujours, la mise en place d'un système automatisé de transport et de stockage (AMHS) n'est devenue essentielle que depuis l'arrivée de la nouvelle génération de wafers de 300mm de diamètre. De plus, les travaux dans la littérature se sont intéressés presque exclusivement aux problèmes liés à la conception du réseau et au choix de la technologie. C'est pourquoi il existe peu de travaux dans la littérature des semi-conducteurs traitant des problèmes de pilotage (e.g. dispatching des véhicules). Ces travaux se basant principalement sur des modèles de simulation *ad hoc*, deux problématiques majeures apparaissent. D'une part, les modèles proposés jusqu'à aujourd'hui étudient des cas industriels précis dont l'objectif est de comparer la performance de certaines règles de gestion de l'AMHS sans aucune démarche d'optimisation. Les conclusions et contributions de ces travaux ne semblent donc pas pouvoir être généralisables. D'autre part, les travaux proposés dans la littérature considèrent le système de transport comme étant une entité isolée du système de fabrication. Sous cette hypothèse, l'impact de l'AMHS sur la performance globale de la fab n'est donc pas étudié.

Compte tenu de ces manques dans la littérature, il semble nécessaire et judicieux de proposer une démarche générale d'optimisation qui traite des problèmes de pilotage opérationnel dans le système de transport énoncés dans la section 3.3 (dispatching et positionnement des véhicules, recherche de tâches). Le couplage de cette démarche avec un modèle de simulation de l'AMHS permettra ensuite une gestion efficace du transport dans la fab.

En prenant en compte ces caractéristiques, notre travail suit les étapes suivantes :

1. Développement d'un modèle conceptuel de l'unité de fabrication à l'aide de méthodologies formelles de modélisation. Cette démarche est essentielle pour la compréhension du système et pour l'identification des entités et leurs interactions (chapitre 4). La description du modèle conceptuel va nous servir par la suite à mieux comprendre, d'une part, où sont situés nos travaux à l'intérieur du système global de supervision de la fab, et d'autre part, à construire le modèle de simulation à événements discrets utilisé pour l'analyse dynamique du système. Selon la terminologie définie dans [42], cette partie de notre travail constitue la *spécification* du modèle lors d'une analyse par simulation. Notre modèle conceptuel nous permet donc de savoir ce que notre système *fait*.
2. Identification, à travers des expériences de simulation, des paramètres critiques et pertinents pour l'optimisation de la performance industrielle et celle du système de transport (chapitre 5). En reprenant les définitions présentées dans [42], cette approche constitue l'*implémentation* du modèle conceptuel, avec laquelle nous allons pouvoir voir *comment* se comporte notre système.
3. Proposition d'une méthodologie d'optimisation. Cette partie du travail est basée sur les résultats obtenus en simulation. L'objectif est d'améliorer les principaux indicateurs de performance de la fab. Notre démarche d'optimisation comporte les étapes suivantes :
   (a) Analyse tactique. La première étape consiste à étudier le problème du point de vue tactique (chapitre 6), où nous devons décider de la "meilleure" façon de gérer les véhicules dans le système (i.e. définition de zones de gestion et distribution des véhicules).
   (b) Gestion opérationnelle. La seconde étape de cette analyse consiste à traiter le problème du point de vue opérationnel (chapitre 7), où nous devons prendre en compte l'évolution dynamique du système et gérer les véhicules à l'aide de stratégies de pilotage opérationnelles. Les résultats de l'optimisation tactique des paramètres sont incorporés dans le modèle de simulation de la fab.

   Notre démarche d'optimisation comporte ainsi les étapes d'analyse tactique (à l'aide de méthodes analytiques) et gestion opérationnelle (à l'aide de la simulation). Le lien entre ces deux approches peut aller dans les deux sens, comme suggéré dans les travaux de Yücesan (e.g. [42, 43]), les approches d'optimisation discrète et de simulation à événements discrets partageant beaucoup d'éléments en commun. Il nous semble donc pertinent de modéliser et résoudre notre problème d'optimisation en utilisant des techniques issues de la Recherche Opérationnelle telles que la programmation linéaire et la simulation. Les solutions obtenues aux niveaux tactique et opérationnel pourront ensuite constituer la base d'une démarche itérative d'aide à la décision.

Chapitre 4

Modélisation conceptuelle de la fab

La modélisation des systèmes de production est difficile en raison du nombre et de la diversité des paramètres à prendre en compte et de la complexité des relations entre ces paramètres. En particulier, dans la fabrication de wafers semi-conducteurs, la complexité de la modélisation du processus de production est due principalement au très grand nombre d'opérations à réaliser sur un même produit, avec des passages multiples sur une même séquence de postes, comme expliqué au chapitre 2.
L'analyse de stratégies de gestion de la production dans les unités microélectroniques nécessite donc impérativement au préalable une bonne modélisation du système de fabrication. Un bon modèle ne sert pas uniquement à faciliter l'utilisation du système de gestion, mais également à suivre de façon efficace le statut des lots et des machines dans les ateliers, ce qui permettra de choisir la bonne stratégie pour le pilotage en temps réel de la production et du transport de lots.
L'objectif de ce chapitre est de conceptualiser, à l'aide de méthodologies formelles de modélisation, tous les composants d'une unité de production de wafers. Premièrement, nous nous focalisons sur le système physique de production, ainsi que sur son système de supervision afin d'identifier les entités du modèle et les interactions entre ces entités. Ensuite, nous modélisons tout le processus de fabrication des wafers, en tenant compte de l'opération du système automatisé de transport. Selon la terminologie définie par Jacobson et Yücesan [42], cette partie de notre travail constitue la *spécification* du modèle lors d'une analyse par simulation. Notre modèle conceptuel nous permet donc de savoir comment notre système est constitué et ce que notre système *fait*.
L'ensemble des travaux présentés dans ce chapitre s'appuie sur des méthodologies formelles de modélisation. En particulier, pour la modélisation du système (physique et de supervision) de production, nous utilisons la Process-Interaction Approach, tandis que pour le processus de fabrication, nous développons un modèle basé sur les réseaux de Petri. La pertinence de l'utilisation de ces deux méthodologies pour la modélisation à des fins de simulation a été discutée par différents auteurs (e.g. [104, 120, 108]). Afin de faciliter la compréhension, avant d'appliquer ces méthodologies à l'unité de production de wafers, nous présentons succinctement les "briques élémentaires" des modèles utilisés.

4.1 Modélisation des systèmes industriels

La notion de système est un terme très ample et général [11, 104]. Les définitions du système sont très variées dans la littérature. Nous pouvons définir un système comme un

"ensemble d'éléments en interaction, organisés en fonction d'un but" [26]. La structure d'un système se compose essentiellement d'une frontière qui le détache de son environnement, de ressources actives et passives hiérarchisées identifiables et dénombrables, et d'un ensemble de relations [34]. D'un point de vue fonctionnel, un système comporte des flux de natures diverses (informations, produits, etc.), des centres de décisions traitant des informations et agissant sur les flux, des boucles de rétroaction informant de ce qui se passe en aval, et des délais de réponse. Il ne faut pas oublier les entrées/sorties matérialisant les rapports de ce système avec son environnement.
Le système auquel nous nous intéressons ici est le système de fabrication de wafers, qui est un système complexe de production à flux discrets. Un système de cette classe évolue de manière discrète dans le temps, c'est-à-dire que les variables du système étudié que l'on désire connaître à tout instant sont discrètes et s'appellent des variables d'état. L'ensemble des valeurs que ces variables peuvent prendre constitue l'espace d'état du système. Il s'ensuit que l'espace d'état est dénombrable et fini. Du fait de la définition de l'espace d'état, chaque changement d'état ou événement se produit de manière discrète dans le temps [66].
Pour appréhender le comportement d'un système et donc détecter et quantifier les conséquences des problèmes liés à leur complexité, on analyse les valeurs prises par les quantités nommées critères de performance. Le terme performance traduit ce que le système est capable d'accomplir : débit du système, taux d'occupation, temps de cycle de fabrication, etc. La mesure de la performance d'un système de production n'est pas toujours évidente et révèle souvent un problème. Dans le cas où on ne peut pas (ou on ne souhaite pas) déterminer les valeurs des critères de performance à l'aide de mesures, on construit un modèle afin de les évaluer. Cette technique est valable que le système de production existe ou non. En effet, évaluer signifie déterminer une quantité par le calcul sans avoir recours à une mesure directe [34]. L'évaluation des performances est donc toujours effectuée à l'aide d'un modèle.
Les modèles sont donc utilisés pour (i) analyser le système et le comprendre, (ii) communiquer avec les autres à propos du système, et (iii) réaliser des expériences [3]. Les définitions du terme modèle sont aussi variées que les définitions du système. Minsky [80] propose la définition suivante d'un modèle :

> "Pour un observateur A, M est un modèle du système S si A peut apprendre, à partir de M, quelque chose d'utile sur le fonctionnement de S".

Un modèle peut également être considéré comme le résultat d'une transcription ou représentation abstraite d'une réalité concrète quels qu'en soient la forme et/ou le langage utilisé.
En pratique, la construction d'un modèle ou modélisation s'effectue soit à partir du système réel s'il existe, soit à partir d'une description de la manière dont il va fonctionner (cahier des charges) s'il n'existe pas encore. On parlera alors de modélisation *a posteriori* si le système existe et de modélisation *a priori* si le système n'existe pas. La construction d'un modèle d'un système existant ou non n'est pas évidente. Afin d'aider ce processus de conceptualisation, plusieurs formalismes existent dans la littérature : ASCI, Conical Methodology, réseaux de Petri, IDEF, DEVS,... [96, 48, 130, 95]. Selon les symboles utilisés par le langage, les formalismes de modélisation peuvent être classifiés en langages ioniques et langages mathématiques [3]. Les premiers concernent l'utilisation de graphismes pour représenter les différents aspects du système qui peuvent ainsi être exprimés plus facilement par des dessins. Les modèles mathématiques utilisent des symboles pour représenter chiffres, variables et autres types d'entités et leurs relations.

Dans notre travail, nous utiliserons d'abord (dans ce chapitre) une représentation graphique pour la modélisation de l'unité de fabrication de wafers. Notre objectif principal est d'identifier les différentes entités du système industriel et les relations entre ces entités. Plus tard au chapitre 6, nous utiliserons des modèles mathématiques lors de l'optimisation du système.

4.2 Modèle du système de fabrication

4.2.1 Méthodologie : La "Process-Interaction Approach"

La Process-Interaction Approach a été conçue par une équipe de chercheurs dirigée par le Prof. Dr. J.E. Rooda à l'Université de Technologie d'Eindhoven, au Pays-Bas [97, 104]. Le principal objectif de cette approche est de conceptualiser les notions relativement vagues des systèmes industriels et des outils de production dans l'optique d'une application bien précise de modélisation et d'analyse des systèmes complexes de production. L'avantage le plus important de cette méthodologie est de prendre en compte à la fois le système physique et le système de supervision de la production, ainsi que leurs interactions. Dans une représentation graphique du système étudié, les cercles représentent les objets ou entités (appelés "process") du système, tandis que les flèches sont le symbole des interactions que développent entre eux les process (informations, ordres, transport de matériaux ou d'énergie, etc.). Chaque process peut ainsi être divisé en plusieurs autres process jusqu'au moment où il n'est plus possible de descendre dans la hiérarchie parce que l'on atteint les process élémentaires ou indivisibles. Cette méthodologie a déjà été appliquée dans les semi-conducteurs pour la modélisation d'une fab à l'étape de conception [120] et à l'analyse focalisée du système de pilotage de machines type cluster [2]. Comme indiqué précédemment, nous nous intéressons ici aux opérations de l'intégralité de l'unité de fabrication : système physique et système de supervision [86].

4.2.2 Modèle du système physique de production

L'environnement dans lequel l'entreprise opère est appelé le marché. Le marché comporte des donneurs d'ordres (ou clients) et des fournisseurs (voir figure 4.1). Le modèle du marché contient ainsi un ensemble de clients (*Clients*), un ensemble de fournisseurs (*Fournisseurs*) et une unité de fabrication de wafers semi-conducteurs (*Fab*). L'unité de fabrication interagit également avec l'environnement naturel. L'unité de fabrication se comporte à la fois comme un donneur d'ordres pour ses fournisseurs et comme un fournisseur pour clients. Elle reçoit des ordres de produits finis (wafers) venant de ses clients et envoie des ordres de matières premières à ses fournisseurs. Ces derniers délivrent ainsi ces matières premières (wafers vierges, produits chimiques, énergie, etc.), qui sont ensuite transformées par l'unité de fabrication afin de livrer le produit fini (wafers) aux clients. Ceci est le niveau le plus agrégé.

En regardant plus en détail, l'unité de fabrication (*Fab* dans la figure 4.1) peut être décomposée en un ensemble de ressources de production (*FabResources*) et un système de supervision et de pilotage de la fabrication (*FabController*), comme illustré dans la figure 4.2. Le système physique de production (*FabResources*) reçoit des messages de son système de supervision (*FabController*), qui envoie des tâches à réaliser au système de fabrication

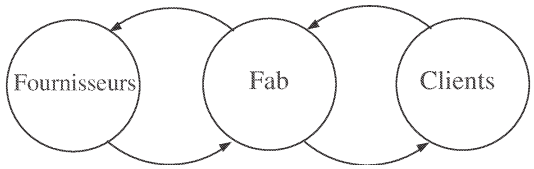


FIG. 4.1 – Environnement de l'unité de fabrication

et coordonne l'approvisionnement des matériaux nécessaires à la fabrication. Le système de production envoie des signaux au système de supervision concernant les états de l'équipement de fabrication (i.e. achèvement des tâches, disponibilité des machines). Étant donné que la capacité de fabrication est fixe, le superviseur de production doit s'assurer que les ressources ne sont ni surchargées ni sous-utilisées. Ceci est réalisé grâce à un suivi régulier du système à l'aide d'un ensemble d'indicateurs de performance appropriés. Dans la fabrication des wafers semi-conducteurs, les indicateurs le plus souvent utilisés sont le temps de cycle et sa variance, le niveau des en-cours, le taux d'utilisation des machines, entre autres.

Le système physique de fabrication (*FabResources*) peut être décomposé comme illustré

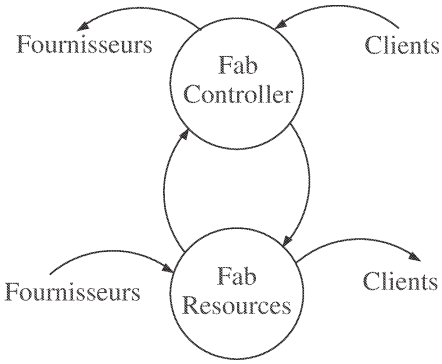


FIG. 4.2 – Modèle du système de fabrication (*Fab*)

dans la figure 4.3. Celui-ci est constitué d'un ensemble d'étagères pour le stockage des lots de wafers en cours de fabrication (*FabStorage*), d'un ensemble de machines usuellement regroupées par familles (*WorkCenters*), et d'un système automatisé de transport (*FabTransporters*). Chaque process *WorkCenters* est composé de plusieurs machines identiques qualifiées pour exécuter les opérations. Chaque groupe de machines (*WorkCenters*) correspond à chacune des familles (i.e. photolithographie, gravure, etc.). À l'intérieur de ces groupes, les machines peuvent être catégorisées en plusieurs types selon leur fonctionnalité. Dans notre modèle, nous distinguons les machines à traitement par batch et les machines séquentielles. Les premières correspondent aux machines capables d'exécuter plusieurs lots à la fois, tandis que les deuxièmes sont des machines qui ne peuvent traiter qu'un lot à la fois. Le process *FabTransporters* est nécessaire pour le transfert des lots entre les opérations exécutées sur les machines (i.e. pour transporter les lots entre les différents familles *WorkCenters*). Il est ainsi composé d'une flotte de véhicules et d'un système de pilotage (voir figure 4.4). Le routage des lots à travers le système (physique) de production ne fait

pas partie du processus de fabrication, et en conséquence, n'ajoute pas de valeur au produit final (mais sans lui, aucune valeur n'est ajoutée). La destination du produit en cours de fabrication est déterminée par le superviseur de production (*FabController*). De cette façon, dans une unité de fabrication automatisée, c'est le superviseur de production qui décide quelle ressource exécute quelle opération et qui suit l'avancement de la production en temps réel.

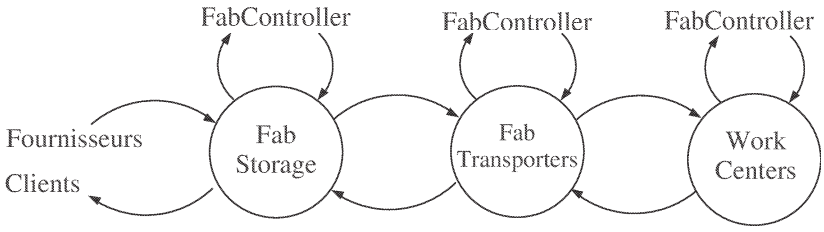


FIG. 4.3 – Modèle *FabResources*

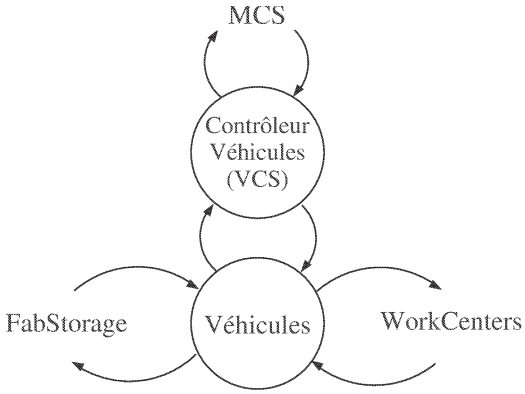


FIG. 4.4 – Modèle *FabTransporters*

4.2.3 Modèle du système de supervision de la fab

Le modèle du système de supervision de production (*FabController*) est illustré par la figure 4.5. Il est divisé en plusieurs sous-systèmes selon les types de fonctions qui doivent être réalisées. Certaines fonctions existent en tant que sous-systèmes individuels, tandis que d'autres sont incorporées dans plus d'un sous-système. Le superviseur doit coordonner les activités des différentes ressources dans l'unité de fabrication et donc prend des décisions concernant les actions qui doivent être réalisées par ces dernières. Les tâches du superviseur sont la planification, l'ordonnancement et dispatching, et le suivi de la production. Les difficultés associées à la planification et à l'ordonnancement/dispatching sont de nature combinatoire et sont la cause des écarts entre les résultats prévus et les résultats réels obtenus par le système de fabrication [109].

Dans la littérature, plusieurs définitions existent pour la tâche de planification [51, 4]. Nous

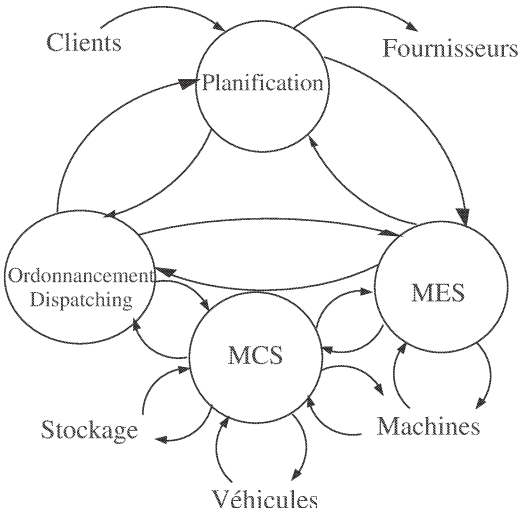


FIG. 4.5 – Modèle du système de supervision (*FabController*)

pouvons considérer que la planification est divisée en sous-tâches telles que la planification de la capacité et du processus de production, les achats et le suivi [109]. La planification de la capacité assure un niveau d'utilisation des ressources pour la fabrication plus ou moins constant au cours du temps, et décide si un ordre de fabrication peut être accepté ou non. La planification du processus de production définit la façon dont les matières premières sont transformées tout au long de la fabrication. La fonction achat gère la commande de ces matières premières aux fournisseurs. Enfin, la fonction de suivi permet de déterminer la prochaine opération qui doit être réalisée sur une pièce à partir du plan de fabrication et de l'information concernant le statut du produit.
L'ordonnancement de la production consiste à définir une date de début et un date d'achèvement pour chaque opération devant être réalisée sur les produits, ainsi que la spécification des ressources qui doivent être utilisées. L'allocation des ressources dépend de la gamme de fabrication du produit, de la qualification des machines pour exécuter les opérations et de la capacité de production disponible dans le système. Le dispatching concerne le suivi en temps réel de l'état des ressources et de la fabrication.
Deux composants très importants à l'intérieur du *FabController* sont le MES (Manufacturing Execution System) et le MCS (Material Control System). Le MES contrôle tous les processus de l'unité de fabrication, y compris le niveau d'en-cours et le statut des ressources en s'appuyant sur une base de données, tandis que le MCS contrôle le système automatique pour le transport des lots. En effet, le MES a les informations concernant la route à suivre par les lots et les instants auxquels les lots doivent être traités sur les machines. Le MCS utilise ensuite ces informations et détermine les destinations et mouvements des lots dans le système. De cette façon, le lancement de nouveaux lots et l'ordonnancement des machines est possible grâce aux informations contenues dans leurs bases de données, ce qui permet aussi la gestion du niveau des encours.
Le transport des lots dans le système peut être réalisé de façon manuelle ou automatique. Puisque nous étudions une unité de fabrication automatisée, notre modèle part donc de l'hypothèse que le transport des lots est réalisé par un système automatique. La gestion de ce système (i.e. le routage et l'affectation des chariots) est effectuée par le VCS (Vehicle

Control System).
Comme nous avons pu remarquer, il existe différentes hiérarchies dans le modèle. Toutes les décisions pour la gestion du système de production doivent être prises par le système de supervision (*FabController*). Puisque le système de production est usuellement géré de façon hiérarchisée, l'algorithme utilisé pour sa gestion est lui aussi de type hiérarchisé. Cette hiérarchie et les relations entre les différents composants du système de gestion nous permettent de déterminer les paramètres et les variables à considérer pour évaluer la performance du système lors d'une étude plus fine des règles de gestion du système automatique de transport des lots ou pour l'ordonnancement de la production.

4.3 Modélisation du processus de fabrication

Une fois identifiés les composants principaux du système physique et de supervision de la fab, nous pouvons nous focaliser sur la compréhension du processus de fabrication lui-même.

4.3.1 Une approche par les réseaux de Petri

Le concept de réseau de Petri (RdP) a été fortement utilisé pour la modélisation et l'analyse des systèmes de production, comme les ateliers flexibles et très récemment pour l'étude de certaines étapes de la fabrication de circuits intégrés [92, 86]. La représentation graphique des RdP permet de visualiser les activités dynamiques des systèmes complexes (e.g. la fabrication de wafers). Cette visualisation est réalisée à l'aide de jetons qui sont introduits dans le réseau. À partir de ce réseau graphique, nous pouvons décrire le comportement du système.
Dans la définition de RdP ordinaires, les notions de temps et de diversité de jetons ne sont pas prises en compte. Si le temps est considéré, l'évolution d'un RdP temporisé ne dépend pas uniquement de l'état du réseau, mais aussi de l'état de l'environnement. Un RdP temporisé est ainsi utile pour l'évaluation des performances du système modélisé. Dans les RdP colorés un identificateur (couleur) est associé à chaque jeton d'une place. Chaque transition peut être franchie de différentes manières représentées par les différentes couleurs de franchissement associées à la transition.

Pour la modélisation du processus de fabrication de wafers, nous utilisons une représentation basée sur un réseau de Petri coloré et temporisé (RdPCT). Certains auteurs ont utilisé les réseaux de Petri pour modéliser certaines étapes du processus de fabrication de wafers (e.g. gravure |45, 74|, test |126|, diffusion |72|). Becker |10| propose un réseau de Petri très complexe pour modéliser tout le processus de fabrication de wafers. Les différents équipements et étapes du processus de fabrication sont considérés dans son modèle. Les activités de transport, cependant, sont modélisées de façon très simpliste uniquement à l'aide d'une transition temporisée stochastique. De plus, comme le dit l'auteur lui-même, la taille du réseau est très importante et donc pas vraiment utile lorsqu'on désire effectuer une analyse des performances. En définisant un type particulier de places et de transitions, Chen *et al.* [19] ont simplifié la modélisation en réseaux de Petri de toute une fab. Dans leur modèle, cependant, les contraintes associées au transport de wafers ne sont pas considérées. D'autres auteurs ont tenté de modéliser le système de transport de façon

détaillée [59, 67], mais sans prendre en compte l'équipement de fabrication. Nous présentons ici une extension du modèle proposé dans [19], dans lequel nous intégrons à la fois toutes les étapes de fabrication et les activités associées au transport des wafers dans la fab.

Formellement, un RdPCT est en fait un graphe biparti défini par un 6-uplet $RdPCT = (P, T, C, I, O, M)$, où :

$P = Pi \cup Pc$ est l'ensemble des places, avec Pi et Pc les ensembles des places ordinaires et des places communicantes, respectivement ;
$T = Tu \cup Tt \cup Ts \cup Tp \cup Tm$ est l'ensemble de transitions, avec Tu l'ensemble de transitions ordinaires, Tt et Ts, respectivement, les ensembles de transitions temporisées déterministes et stochastiques, Tp l'ensemble de transitions de registre des couleurs, Tm l'ensemble de macro-transitions (i.e. l'agrégation d'un processus plus complexe) ;
C est la description des couleurs ;
I est l'ensemble des arcs d'entrée aux places et aux transitions ;
O est l'ensemble des arcs de sortie des places et des transitions ;
M est le vecteur de marquage initial
$P \cap T = \emptyset$ et $P \cup T \neq \emptyset$.

La représentation graphique utilisée pour notre RdPCT est illustrée dans la figure 4.6.

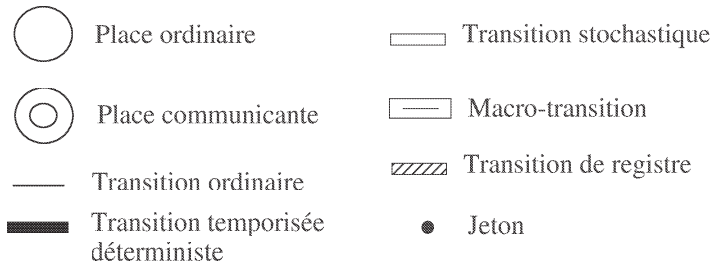


FIG. 4.6 – Notation graphique du RdPCT

4.3.2 Le modèle

Notre modèle est une extension du modèle proposé dans [19], qui est focalisé uniquement sur l'ordonnancement de la production. Ici, nous considérons explicitement tous les composants du système de transport des lots et l'ajoutons à la modélisation. Chaque lot qui entre dans la salle blanche a une séquence d'opérations prédéfinie par la technologie associée au type de produit qui doit être livré au client. Si nous ne faisons pas allusion à la couleur associée aux places ou transitions, alors la couleur de ces places p et ces transitions t est $C(p) = C(t) = \{xzws \mid x \in X, z \in Z, w \in W, s \in S\}$, où la couleur est codée comme une chaîne de caractères de la forme $xzws$, où X représente les gammes de fabrication, Z est le code du lot, W représente l'étape du processus de fabrication dans laquelle se trouve le lot de wafers, et S représente le type d'opération à réaliser. De cette façon, l'unité de production a été divisée en n familles de machines selon le type d'opération qu'elles sont qualifiées à exécuter (voir figure 4.7). À l'instant initial, un jeton représentant un

lot entre dans le système avec une couleur $xzws_0$. Ensuite, ce jeton est marqué avec la couleur $xzws_1$ signifiant que la première opération à réaliser sera effectuée sur la famille de machines assignée dans la place *dispatchinglot* conformément à la route qu'il doit suivre dans le système. Avant d'arriver à la machine pour l'exécution de l'opération, le lot est transporté par le système automatique (macro-transition *amhs*). Une fois l'opération achevée, la couleur du lot est incrémentée d'une unité et le lot transporté à la famille de machines qualifiée pour exécuter l'opération suivante. Lorsque le produit est terminé, le jeton (lot) passe dans la place ordinaire *end*. Cette description est représentée dans la figure 4.7. Les places et transitions figurant dans ce modèle ont les significations suivantes :

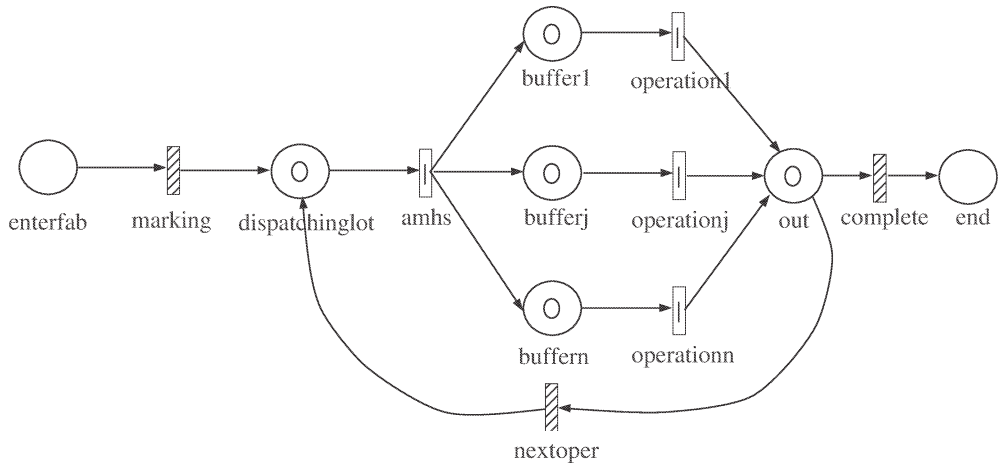


FIG. 4.7 – Modèle général du processus de fabrication

- *enterfab* est une place ordinaire. Un jeton dans cette place représente un lot de plaquettes vierges qui entre dans le système avec une couleur spécifique selon la route de fabrication associée au type de produit à fabriquer (i.e. $xzws_0$).
- *marking* est une transition de registre de couleurs. Lorsque cette transition est franchie, le jeton est marqué avec la couleur s qui correspond à la toute première opération à réaliser sur le wafer selon la gamme qu'il doit suivre.
- *dispatchinglot* est une place communicante qui peut être vue comme le système d'ordonnancement de la fabrication. Un jeton dans cette place doit être dirigé vers l'une des familles de machines selon l'opération à effectuer (i.e. la couleur du jeton) et la qualification des machines pour réaliser cette opération. Ce système a été décrit dans la section 4.2.3.
- *amhs* est une macro-transition qui correspond au transfert automatique des lots entre les opérations de fabrication. Cette transition sera présentée plus en détail par la suite.
- $buffer_j$ $(j = 1, 2, ..., n)$ est une place communicante. Elle représente la liste d'ordonnancement. Un jeton dans cette place est un lot attendant son traitement devant une famille de machines. Cette place contient aussi les règles de choix si plusieurs machines dans la famille sont disponibles pour exécuter une opération donnée. Ces règles sont principalement basées sur les priorités d'exécution des lots, i.e. FIFO, SPT, EDD, critical ratio (CR).
- $operation_j$ $(j = 1, 2, ..., n)$ est une macro-transition qui représente le processus d'exécution d'une opération sur le wafer. Cette transition sera analysée par la suite.

- *out* est une place communicante. Un jeton dans cette place est un lot qui a fini le traitement de l'opération en exécution. Le jeton peut ainsi franchir l'une des deux transitions et aller soit à *nextoper*, soit à *complete*, selon qu'il ait à réaliser une nouvelle opération ou qu'il ait fini son cycle de production, respectivement.
- *nextoper* est une transition de registre. Lorsqu'elle est franchie, la couleur du jeton est incrémentée d'une unité. Elle représente l'achèvement d'une opération sur un lot (jeton) qui passe maintenant à son opération suivante.
- *complete* est une transition de registre. Elle est franchie lorsqu'un lot (jeton) a fini toutes les opérations de sa gamme opératoire et passe ainsi à la place *end*.
- *end* est une place. Un jeton dans cette place veut dire que toutes les opérations nécessaires à la fabrication du produit souhaité ont été accomplies et le produit peut être livré au client.

Le transport des lots

Nous présentons maintenant le modèle du processus de transfert des lots de wafers qui correspond à la macro-transition *amhs* dans la figure 4.7. Les places et transitions du modèle présenté dans la figure 4.8 ont les significations suivantes.

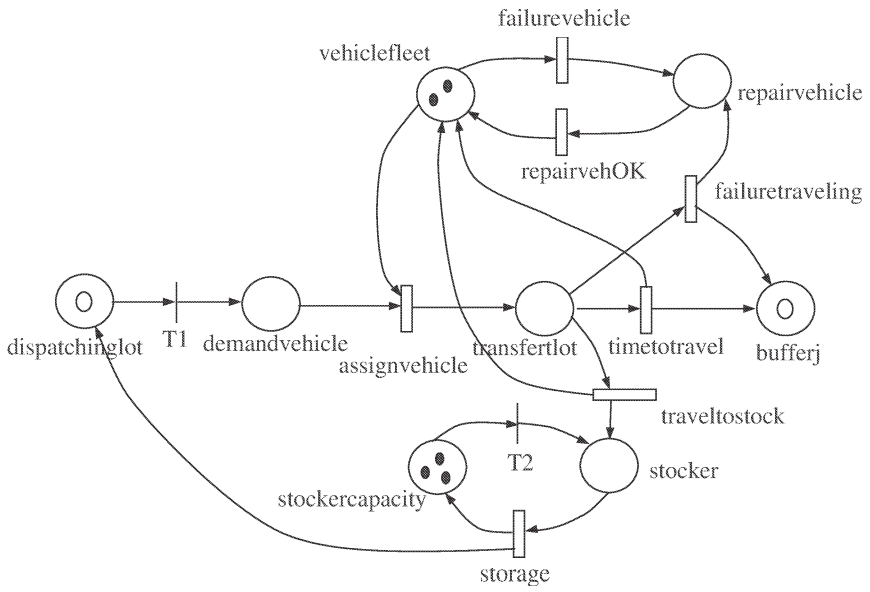


FIG. 4.8 – Modèle du transport des lots

- *demandvehicle* est une place ordinaire. Un jeton dans cette place est un lot qui demande un véhicule pour effectuer le service de transfert.
- *assignvehicle* est une transition stochastique. Son franchissement est possible s'il y a des véhicules disponibles pour le transfert et si un lot en a fait la demande. Son temps de franchissement dépend du temps d'attente pour l'arrivée d'un véhicule et du temps nécessaire au système de pilotage pour réaliser l'affectation d'un véhicule selon les règles définies dans le système, comme expliqué dans la section 4.2.3.
- *transfertlot* est une place ordinaire. Un jeton dans cette place est un lot qui est transporté. Selon la destination finale (couleur du jeton), le lot peut être transféré soit vers une

étagère pour stockage temporaire (la transition *traveltostock* sera ensuite franchie), soit vers une machine (la transition *timetotravel* sera franchie). En cas de panne du véhicule, le transfert est interrompu est un jeton est dirigé vers la transition $failuretraveling$.
- *timetotravel* est une transition stochastique. Le temps de franchissement dépend du temps nécessaire au transfert du lot entre sa localisation actuelle et sa destination finale (la machine).
- *traveltostock* est aussi une transition stochastique et son temps de franchissement dépend du temps nécessaire pour le transfert du lot vers la zone de stockage temporaire.
- *stocker* est une place ordinaire. Un jeton dans cette place est un lot stocké de façon temporaire dans une étagère (stocker).
- *stockercapacity* est une place ordinaire. Elle est marquée à l'instant initial car elle représente la capacité de stockage temporaire de lots.
- *storage* est une transition stochastique correspondant au temps qu'un lot a passé en stockage avant d'être transféré vers une opération. Le franchissement de cette transition génère deux jetons. Le premier est le lot qui va maintenant vers la place *dispatchinglot* afin qu'il puisse être dirigé vers sa prochaine opération. Le deuxième jeton correspond à la place dans le stocker qui vient d'être libérée par le lot et se dirige vers la place $stockercapacity$.
- $vehiclefleet$ est une place ordinaire marquée à l'instant initial. Le nombre de jetons dans cette place représente le nombre de véhicules disponibles pour effectuer les opérations de transfert.
- $repairvehOK$, $failurevehicle$ et $failuretraveling$ sont des transitions temporisées stochastiques. Le temps de franchissement de $repairvehOK$ dépend du temps moyen pour la réparation du véhicule. Les temps de franchissement de $failurevehicle$ et $failuretraveling$ dépendent du temps moyen entre deux pannes. Si la transition $failuretraveling$ est franchie, cela veut dire que la panne du véhicule s'est produite lorsqu'il transportait un lot. Dans ce cas, le véhicule est enlevé du réseau (pour éviter le blocage du tronçon) et le lot est placé dans la file d'attente de la machine (place $buffer_j$).
- *repairvehicle* est une place ordinaire. Les jetons dans cette place représentent des véhicules en réparation.

Exécution de lots sur les machines

Le RdPCT pour le traitement des lots sur les machines est illustré dans la figure 4.9 ci-après. Cette représentation correspond à la macro-transition $operation_j$ de la figure 4.7. Les significations des places et transitions sont les suivantes :
- $enter_j$ $(j = 1, ..., n)$ est une transition ordinaire qui est franchie lorsqu'un lot (jeton) doit commencer son traitement sur une machine.
- $betimecritical_j$ $(j = 1, ..., n)$ est une place ordinaire. Un jeton dans cette place (avec n'importe quelle couleur) correspond à un lot avec une priorité d'exécution sur la machine plus importante car il va bientôt dépasser une contrainte de traitement. Un jeton dans cette place empêche le franchissement de la transition $enter_j$.
- $timecritical_j$ $(j = 1, ..., n)$ est une transition temporisée déterministe qui sera franchie par un jeton se trouvant dans la place $buffer_j$ et qui a dépassé le délai d'attente pour l'exécution d'une opération. Le temps nécessaire au franchissement de cette transition dépend de la couleur du jeton (i.e. si l'opération à exécuter nécessite une attente ou ne peut pas excéder une limité de temps).
- $entercrit_j$ est une transition ordinaire. Elle est franchie si et seulement si il y a une

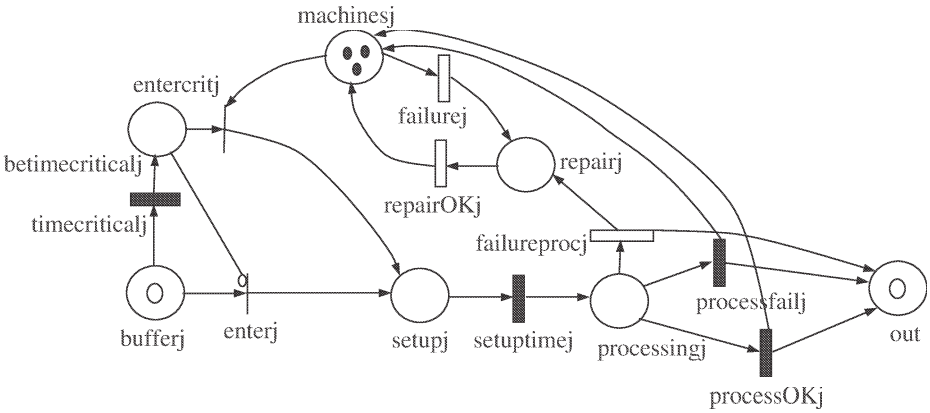


FIG. 4.9 – Modèle de la fabrication de lots

machine disponible dans $machines_j$ afin qu'un lot se trouvant dans $betimecritical_j$ commence son traitement. Après le franchissement, le jeton sera placé dans $setup_j$.

- $setup_j$ $(j = 1, ..., n)$ est une place. Un jeton dans cette place représente une machine prête à démarrer une opération.
- $setuptime_j$ $(j = 1, ..., n)$ est une transition temporisée déterministe. Son temps de franchissement est donné par le temps de préparation de la machine selon la couleur (i.e. type de produit) du jeton qui va être traité. Ce temps est nul si la machine ne requiert pas de temps de préparation ou si la couleur du jeton arrivant et la couleur du jeton précédent sont les mêmes.
- $processing_j$ $(j = 1, ..., n)$ est une place. Un jeton dans cette place est un lot en exécution sur une machine. Si l'opération exécutée comporte aussi une inspection entraînant un refus du lot, alors le jeton doit franchir la transition $processfail_j$. Si au contraire, le lot est conforme, alors il franchit la transition $processOK_j$.
- $processOK_j$ $(j = 1, ..., n)$ est une transition temporisée déterministe. Le temps de franchissement est une fonction de la couleur (type d'opération) du jeton. Le franchissement représente donc l'achèvement de cette opération et génère deux jetons : le premier représente une machine qui se libère et se dirige à la place $machines_j$, tandis que le deuxième est le lot qui doit maintenant aller dans la place out.
- $processfail_j$ $(j = 1, ..., n)$ est aussi une transition déterministe. De façon similaire, le temps de franchissement de cette transition est fonction de la couleur du jeton. Lorsqu'elle est franchie, deux jetons sont créés : le premier va vers la place $dispatchinglot$ afin de le re-ordonnancer pour une opération de reprise ; et le deuxième représente la libération de la machine et se dirige donc vers la place $machines_j$.
- $machines_j$ $(j = 1, ..., n)$ est une place ordinaire qui est marquée à l'instant initial. Les jetons dans cette place représentent les machines qualifiées disponibles pour effectuer l'opération j.
- $repair_j$ est une place ordinaire. Un jeton dans cette place est une machine qui est en maintenance (préventive ou suite à une panne inattendue ou à un réglage).
- $repairOK_j$, $failure_j$, et $failureproc_j$ $(j = 1, ..., n)$ sont des transitions temporisées stochastiques. Le temps de franchissement de $repairOK$ dépend du temps moyen de réparation (MTTR) de la machine. Les temps de franchissements des transitions $failure_j$ et $failureproc_j$ dépendent du temps moyen entre deux pannes. Le franchissement de

la transition $failureproc_j$ veut dire que la machine est tombée en panne pendant le traitement d'un lot.

Le traitement des batchs

La description de la fabrication sur les machines à traitement par batch est un peu plus complexe, mais sa représentation est la même que celle décrite dans la figure 4.9. La couleur de $enter_j$, $setup_j$, $entercrit_j$, $setuptime_j$, $processing_j$, $failure_j$, $processOK_j$ et $processfail_j$ dépend de la taille du batch. Par exemple, si la taille du batch est deux lots, alors la couleur est $C(t) = C(p) = \{bx_1z_1w_1s_1x_2z_2w_2s_2 \mid b \in W, x_1, x_2 \in X, z_1, z_2 \in Z, w_1, w_2 \in W, s_1, s_2 \in S\}$. La transition $enter_j$ (resp. $entercrit_j$), avec $j = 1, ..., n$, est une transition ordinaire. Un jeton franchissant cette transition représente un batch pour une opération normale (resp. urgente dû à une fenêtre de temps). Il doit y avoir deux jetons dans la place $buffer_j$ (resp. $betimecritical_j$) avec les couleurs $x_1w_1z_1s_1$, $x_1 \in X$, $z_1 \in Z$, $w_1 \in W$, $s_1 \in S$ et $x_2w_2z_2s_2$, $x_2 \in X$, $z_2 \in Z$, $w_2 \in W$, $s_2 \in S$, et un jeton dans la place $machines_j$ avec la couleur $b \in W$. Après franchissement, un seul jeton entrera dans la place $setup_j$ avec comme couleur $bx_1z_1w_1s_1x_2z_2w_2s_2$ ou $bx_2z_2w_2s_2x_1z_1w_1s_1$.

4.4 Conclusion

Dans ce chapitre, nous nous sommes intéressés à la modélisation conceptuelle d'une unité de fabrication de wafers semi-conducteurs. Nous avons utilisé des méthodologies formelles de modélisation pour mieux comprendre aussi bien le système physique de production et son système de supervision que le processus de fabrication lui-même. Ces modèles, qui facilitent à la fois la compréhension du système de gestion et de supervision de la production, constituent une base fondamentale pour le développement d'une étude simulatoire pour l'évaluation des performances. Selon la terminologie définie dans [42], cette partie de notre travail constitue la *spécification* du modèle lors d'une analyse par simulation. Notre modèle conceptuel nous permet donc de savoir ce que notre système *fait*.
Le chapitre suivant présente une analyse plus détaillée des stratégies de gestion, notamment, pour la gestion du système automatique de transport de plaquettes (e.g. règles de dispatching des véhicules). L'objectif principal est de comparer la véritable influence des stratégies de pilotage du système de transport sur les performances de la fabrication afin de déterminer les bases pour une démarche globale d'optimisation.

Chapitre 5

Présentation et analyse du cas industriel

Nos travaux ont été réalisés en partenariat avec STMicroelectronics, une compagnie Franco-Italienne de semi-conducteurs. L'unité de production considérée dans cette étude est une nouvelle fab localisée à Catane (Sicile, Italie) qui produira des wafers de 300mm de diamètre. Cette fab étant complétement automatisée, notre travail s'inscrit dans une volonté de l'entreprise d'identifier des politiques "optimales" pour la gestion du transport de wafers dans la fab, et ainsi mieux comprendre les interactions entre l'AMHS et les ressources de production.
Dans ce chapitre, nous présentons tout d'abord l'unité de fabrication considérée dans cette thèse pour la mise en œuvre des stratégies de pilotage du système de transport automatisé. Nous nous intéressons plus particulièrement au processus de fabrication et au système automatisé de transport. Dans la deuxième partie du chapitre, nous étudions les problématiques liées à la gestion de l'AMHS dans la fab et, en particulier, nous analysons, *via* simulation, l'impact de certaines règles de dispatching des véhicules sur la performance globale de la fab. L'objectif de cette étude de simulation est d'avoir plus d'informations sur le comportement du système de fabrication lors des changements des politiques de gestion du transport interne (e.g. sensibilité des indicateurs de performance, identification des paramètres critiques, comparaison de règles de dispatching). Ce chapitre nous donnera ainsi des éléments de base pour établir notre démarche globale d'optimisation.

5.1 Présentation du cas industriel

5.1.1 Le processus de fabrication et le système de transport

Comme vu précédemment, le *layout* d'une fab est divisé en un certain nombre de *bays* (couloirs) qui contiennent les équipements pour fabriquer les wafers. Cette configuration en *bays* augmente considérablement la circulation des produits, le processus de fabrication étant de type re-entrant [57]. La configuration en *bay* a cependant des avantages logistiques [14], comme expliqué dans la section 2.3.
Dans notre unité de production, il y a plus de 200 machines qui peuvent être regroupées en près de 70 familles, chacune composée d'une ou plusieurs machines parallèles identiques. Les données présentées sont basées sur des données réunies lors de l'étape de conception de la fab. Les temps opératoires dans notre modèle (et dans la fab réelle) incluent les temps

de préparation (*set-ups*). Puisque nous étudions une fab complètement automatisée, les temps d'attente dus à l'absence des opérateurs ne sont pas considérés. Toutes les visites de tous les lots à une famille de machines spécifique ont la même distribution du temps opératoire.
La taille d'un lot de wafers est connue (25 plaquettes par lot) et reste constante tout au long de l'étude. Par rapport au problème de composition et recomposition de lots dans une cassette (FOUP[1]), nous considérons, sans perte de généralité, qu'un FOUP contient un seul lot. La gamme opératoire comporte un total d'environ 700 étapes de fabrication pour un seul produit. Cette gamme représente le flux classique re-entrant qui caractérise la fabrication de semi-conducteurs où chaque lot doit passer plusieurs fois sur le même type de machine à différents stades de sa gamme. En particulier, dans la gamme opératoire considérée dans cette thèse, un lot doit passer 40 fois par l'atelier de photolithographie.

Les opérations de transport considérées dans cette étude peuvent être classées comme inter-bay (entre les ateliers de la fab) et intra-bay (à l'intérieur des *bays*) et sont réalisées par un système unifié de transport automatisé (*unified* AMHS), comme dans la figure 5.1. Le fait que l'AMHS soit unifié implique que le réseau est conçu de telle façon que les véhicules peuvent transporter un lot directement d'une machine vers une autre. Le système est un convoyeur aérien (*Overhead Hoist Transporter*, OHT) avec véhicules capables de prendre le FOUP par la partie supérieure. Le principal avantage d'installer un convoyeur aérien consiste en une meilleure utilisation de l'espace au sol [115] et donc une réduction de la superficie totale de la salle blanche. Contrairement aux systèmes de transport avec chariots autoguidés (AGV), l'alimentation en énergie pour les véhicules d'un convoyeur aérien est faite directement par les rails. La communication entre les véhicules et l'ordinateur central (*Material Control System*, MCS) est également réalisée à travers les rails. Les véhicules ont chacun leur propre moteur. De cette façon, des problèmes souvent retrouvés dans les systèmes avec AGV tels que la gestion de l'alimentation (chargement de la batterie), l'identification d'obstacles ou la transmission des ordres, ne sont pas critiques pour notre cas. L'autre avantage de ce système est la facilité pour déterminer le positionnement exact des véhicules dans le réseau [115].

Dans notre fab, le réseau de transport est un réseau unifié uni-directionnel. Tous les véhicules sont du même type (i.e. ils ont tous les mêmes caractéristiques) et de capacité unitaire (i.e. ils ne peuvent transporter qu'un FOUP à la fois). Le pilotage du système est réalisé par une interaction entre le *Manufacturing Execution System* (MES), le *Material Control System* (MCS) et le système de contrôle de véhicules (*Vehicle Control System*, VCS). Ces systèmes font partie intégrante du système global de supervision de la fab (voir figure 4.5 en page 38) [92]. Le MES controle le mouvement logique des lots entre les opérations de fabrication. L'ordre dans lequel les lots sont exécutés sur les machines dépend de la politique d'ordonnancement implémentée. A ce sujet, remarquons que notre étude est limitée uniquement au système de transport (i.e. le transfert des lots entre les opérations de fabrication) et nous n'analysons pas la performance du système par rapport à des règles d'ordonnancement de la production. Néanmoins, nous prenons en compte cet aspect et intégrons les interactions entre les activités de transport et de production. Le MCS est chargé de convertir les ordres logiques du MES en commandes de mouvements physiques pour le FOUP qui sont envoyées aux équipements (machines), étagères (stockers)

[1]Rappelons que le FOUP est la cassette hermétique qui contient le wafers de 300mm de diamètre.

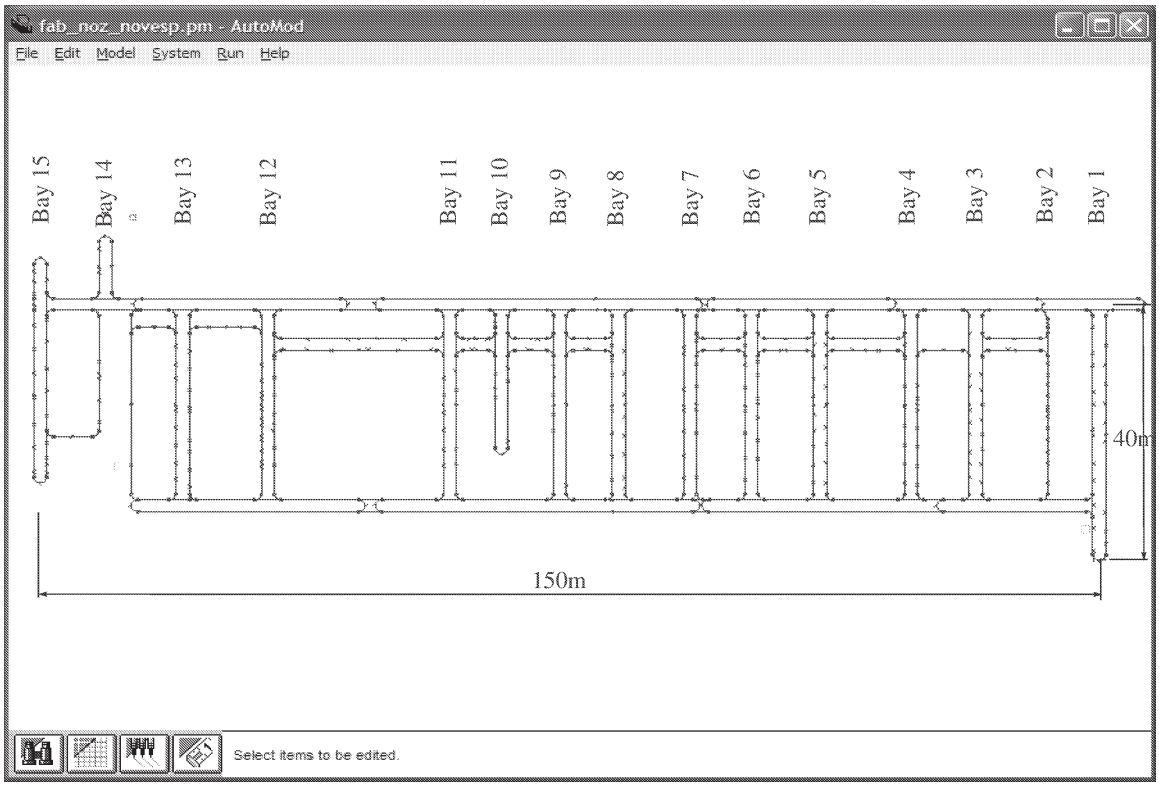


FIG. 5.1 – Configuration du réseau de transport (AMHS)

et véhicules. Après réception d'une commande provenant du MCS, le logiciel de contrôle de véhicules (VCS) exécutera la série d'opérations nécessaires pour prendre, déplacer et délivrer le FOUP vers sa destination finale.

5.1.2 Particularités du cas industriel

Cette section présente les caractéristiques qui distinguent cette fab particulièrement complexe des autres unités de production de wafers étudiées dans la littérature. Plus précisément, la taille de la fab et la complexité intrinsèque de la fabrication de wafers font de cet environnement de production un véritable challenge pour la recherche en génie industriel. Les caractéristiques particulières de notre fab sont :

- **Taille du réseau de transport** : La taille du réseau de transport est la première caractéristique importante qui distingue notre cas des autres configurations des systèmes traditionnellement étudiés dans la littérature. En effet, la superficie réelle de la salle blanche de notre fab est l'équivalent d'un terrain de football. De plus, le nombre de machines nécessaires pour la fabrication, et donc le nombre de points de chargement et de déchargement dans le système est un autre facteur très important car nous considérons plus de 200 points dans le réseau de transport.
- **Système de transport unifié** : Comme indiqué au chapitre 3, jusqu'il y a quelques années, seuls les systèmes **non-unifiés** de transport ont été réellement étudiés dans les fabs. La principale raison de ce choix est fondamentalement basée sur le fait que, dans

un système non-unifié, les véhicules ne pouvant pas sortir de leur boucle sont contraints d'effectuer des transferts soit inter-bay, soit intra-bay [91]. Ainsi, l'implémentation de politiques de gestion et contrôle (i.e. dispatching) est hautement simplifiée, malgré la taille très élevée de la flotte. Dans notre fab, puisque le système est **unifié**, les véhicules sont autorisés à transporter les lots directement d'une machine vers une autre. Ainsi, la spécialisation des véhicules est plus difficile et coûteuse, compte tenu de l'investissement en capital pour qu'ils n'effectuent que soit des transferts inter-bay, soit des transferts intra-bay.

- **Taille de la flotte de véhicules** : Le nombre de véhicules nécessaire dans le système, pouvant facilement atteindre la trentaine (selon la politique de gestion mise en œuvre)[2], est une autre caractéristique spécifique à notre unité de production [87].
- **Nombre d'étapes de fabrication** : Le processus de fabrication de notre fab comporte environ 700 étapes, ce qui implique un plus grand nombre de visites à l'atelier de photolithographie (couches de produits chimiques). Ceci représente plus du double d'étapes de fabrication par rapport aux fabs existantes de l'ancienne génération de wafers (200mm de diamètre).
- **Prise en compte du contrôle de la production dans le modèle de simulation du système de transport** : Nous voulons que notre modèle de simulation représente le plus possible le véritable comportement de la fab. Les intéractions entre la gestion du transport dans la fab et la gestion de la production sont importantes [20, 7, 63, 62]. Cependant, cette recherche ne s'intéresse pas à l'ordonnancement de la production ni à l'étude des dites interactions. Même si la performance de la fab n'est pas analysée par rapport à l'implémentation de diverses politiques de gestion de la production, notre objectif est de prendre en compte cet élément et de l'intégrer lors de l'analyse des stratégies de pilotage de l'AMHS. Ainsi, des règles réalistes pour la gestion de la production sont intégrées dans le modèle de simulation de l'AMHS. Ces règles seront expliquées par la suite dans la section 5.2.1.

5.2 Etude de simulation

5.2.1 Présentation succincte du modèle

Le modèle de simulation a été implémenté sur un PC Pentium® 4 (3,4 GHz) en utilisant le progiciel AutoMod™ [3]. Les principaux éléments du modèle sont décrits en détail dans [103] et ont été définis grâce aux modèles conceptuels présentés au chapitre 4. Dans notre modèle, l'entité principale est la cassette ou FOUP. Cette cassette contient 25 wafers et correspond à un lot de production. De cette façon, pour nous référer à l'entité, nous parlerons dans la suite indistinctement de lot, FOUP ou cassette.
Les ressources modélisées sont les machines de production, les véhicules et les étagères pour le stockage intermédiaire des lots. Les machines ont été modélisées de façon indépendantes et sont localisées physiquement dans la fab. Chaque opération de la gamme est réalisée sur

[2] Comme nous avons expliqué dans la section 5.1.1, notre étude ne considère que les lots de production. Dans la fab réelle, nous avons aussi des lots dits NPW (*Non Product Wafers*) qui servent à la qualification des machines et aux tests suite aux pannes. Dans les conditions réelles de fabrication, la taille de la flotte peut très facilement atteindre 60 ou 70 véhicules dans le réseau. Comme nous verrons ensuite dans la section 5.2.1, notre étude est réalisé à laide d'un modèle de la fab qui ne considère pas les lots NPW car la gestion du système devient plus complexe

[3] AutoMod™ est une marque déposée de Brooks Automation, Inc. - AutoSimulations Division.

une famille de machine, et chaque famille est composée d'un certain nombre de machines parallèles identiques. Les étagères pour le stockage intermédiaire des cassettes (appelées stockers) sont aussi prises en compte par notre modèle avec une capacité illimitée de stockage. Leur localisation physique dans la fab est également connue. Les véhicules sont les ressources dédiées au transfert des FOUPs entre les opérations. Les déplacements autorisés pour transporter une entité sont exclusivement soit d'une machine vers un stocker, soit d'un stocker vers une machine. Dans ce modèle, nous n'acceptons pas la livraison directe d'une machine vers une autre machine (*tool to tool delivery*). Tous les véhicules sont considérés identiques (i.e. ils ont la même vitesse) et de capacité unitaire (i.e. ils ne peuvent transporter qu'une entité à la fois).
La gestion des files d'attente est un autre élément important de notre modèle de simulation. Pour chaque machine, une file d'attente à l'arrivée et à la sortie est modélisée. Ces files ont une capacité limitée donnée par le nombre de places (*loadports*) disponibles pour le chargement et le déchargement des FOUPs. En raison de cette capacité limitée des queues devant les machines, le dispatching des lots vers les machines se fait en flux tiré : les lots doivent passer obligatoirement par une étagère de stockage intermédiaire (stocker) où ils attendent qu'une place se libère devant la machine, et c'est à cet instant que la machine appelle un lot du stocker. Lorsqu'un lot est appelé par une machine, il émet une requête de transport et attend l'arrivée d'un véhicule. La file d'attente est gérée en utilisant la règle FIFO (*First In First Out*), ce qui implique une même priorité pour tous les lots arrivant dans la queue[4], leur place dans la liste d'exécution étant donné uniquement par l'ordre d'arrivée dans la file). De manière similaire, un lot terminé sur une machine, émet une demande de transport pour aller vers un stocker.
Dans notre modèle de simulation, nous avons considéré une seule gamme opératoire. Tous les lots doivent donc suivre la même séquence d'opérations. Dans la réalité, il peut y avoir une grande variété de produits dans la fab, chacun suivant sa propre gamme opératoire. Le fait d'avoir la même séquence pour tous les lots n'est pas une perte de généralité pour étudier la performance de l'AMHS car cette gamme représente bien la complexité du processus de fabrication, avec environ 700 étapes de fabrication et de multiples passages sur une même séquence de postes (flux re-entrant caractéristiques de la fabrication de semiconducteurs). Aussi, seuls les lots de production sont étudiés. Les wafers nécessaires pour les tests et qualification des machines (dits NPW, *Non Product Wafers*) ne sont pas pris en compte dans notre étude car la gestion du système deviendrait plus complexe sans vraiment ajouter de la valeur à notre analyse.

5.2.2 Règles de pilotage du système de transport

Dans cette section, nous décrivons en détail les stratégies mises en place dans le modèle de simulation de la fab qui ont été choisies pour répondre à chacun des problèmes liés au pilotage du système de transport.

[4]De manière générale dans la fabrication de wafers, nous pouvons trouver des lots dits "normaux" (i.e. de lots qui ne sont pas urgents pour le traitement à l'arrivée dans la file d'attente devant les machines), et des lots dits "urgents", qui seront toujours exécutés avant les lots "normaux". La définition des niveaux d'urgence, et la méthode pour quantifier la priorité, peuvent varier d'un fabricant à un autre, et d'une fab à une autre. De façon générale, les différents niveaux de priorité des lots "urgents", par ordre croissant, sont : *hot lots*, *superhot lots* et *rush*. Aussi, les lots "normaux" peuvent avoir différents niveaux de priorité pour leur exécution, mais, d'un point de vue conceptuel, cela relève plutôt de la politique de dispatching de production (i.e. gestio des files d'attente) mise en place.

Dispatching des véhicules

Comme expliqué au chapitre 3, le problème de dispatching de véhicules consiste à trouver un ordonnancement permettant de traiter toutes les demandes de transport. L'objectif est donc de sélectionner une demande parmi toutes celles qui attendent un transfert et de l'affecter à l'un des véhicules disponibles. Nous verrons par la suite les règles mises en place dans notre modèle.

- **Sélection d'une tâche** : Comme expliqué dans la section 3.3.1, le problème de sélection d'une tâche (*task selection problem*) se présente lorsqu'il y a un seul véhicule libre et que plusieurs demandes ont été émises. On a alors des politiques de dispatching dites à l'initiative du véhicule. Dans cette étude, notre intérêt est de comparer la performance de trois règles en particulier : FCFS (*First Come First Serve*), STD (*Shortest Travel Distance*) et une règle hybride combinant les deux précédentes et que nous appellerons CL-FCFS (*Current Location-First Come First Serve*). Nous rappelons brièvement les définitions de ces règles de dispatching. Pour mémoire, FCFS est la règle utilisée dans la littérature des semi-conducteurs et par les systèmes implémentés dans les fabs.
  * FCFS : Le service de transport se fait par ordre chronologique des demandes.
  * STD : Lorsqu'un véhicule se libère, les distances le séparant des stations demandeuses sont évaluées. Le véhicule choisi est celui dont la distance à la station à servir est la plus faible.
  * CL-FCFS : Lorsqu'un véhicule est disponible pour satisfaire une demande de transport, il regarde d'abord si la station actuelle est demandeuse. Si c'est bien le cas, cette requête est servie, sinon, les demandes sont satisfaites selon la règle FCFS précédemment expliquée.
- **Sélection d'un véhicule** : Comme indiqué au chapitre 3, le problème de sélection d'un véhicule (*vehicle selection problem*) se présente lorsqu'un seul poste de travail émet des requêtes et qu'il y a plus d'un véhicule disponible. Dans cette étude, nous avons choisi la règle LIV (*Longest Idle Vehicle*). Cette règle choisit toujours le véhicule qui est resté inactif le plus longtemps. Rappelons que la règle LIV permet d'équilibrer la charge de travail pour tous les véhicules dans le système.

Routage des véhicules et gestion du trafic

Pour le routage des véhicules, nous avons choisi le chemin le plus court. Après quelques simulations préliminaires [87, 88], nous avons constaté que le chemin le plus court est aussi le plus rapide, compte tenu de la configuration et des caractéristiques du réseau. Par conséquent, le problème de routage n'est pas pertinent pour notre étude, comme nous le verrons également lors de l'analyse des résultats.
En ce qui concerne la gestion du trafic, une distance de sécurité a été déterminée pour éviter les collisions entre les véhicules. Lorsqu'un véhicule détecte un obstacle (un autre véhicule) sur son chemin, alors il ralentit ou, le cas échéant, s'arrête complètement afin de pouvoir respecter cette distance de sécurité. Le même phénomène se produit aux intersections. De cette façon, l'animation et les résultats de ces simulations ont montré que les problèmes liés à la congestion du réseau ne se posent pas compte tenu également de la taille et de la configuration du système [87, 82] par rapport au nombre de véhicules.

Définition de zones dans la fab

Afin de mieux gérer le système de transport et de satisfaire les demandes de façon efficace, nous allons tester l'implémentation d'une gestion des véhicules et des demandes de transport par zones dans la fab. Nous allons donc comparer expérimentalement deux politiques : une sans diviser la fab en zones (autrement dit, avec une seule zone) et une autre avec plus d'une zone. La gestion opérationnelle des véhicules est expliquée par la suite.

Politique "1-zone" : Cette politique de gestion de l'AMHS constitue la politique de base du système. Elle est en fait utilisée dans les travaux de la littérature. Les véhicules sont autorisés à servir toutes les demandes de transport provenant de tous les points dans le réseau, selon la règle de dispatching mise en place. Ainsi, la performance du système de transport dépend fortement du nombre de véhicules dans le système et de la règle de dispatching implémentée. Nous appellerons cette politique "1-zone", car la fab est considérée comme une seule zone de gestion.

Politique "3-zone" : Avec cette politique, la fab est divisée en 3 zones de gestion. Chaque zone est composée d'un ensemble de points choisis selon leur localisation géographique dans le réseau. Les véhicules peuvent se déplacer dans tout le réseau, mais lorsqu'un véhicule entre dans une zone, il ne peut servir que les demandes émises par les points appartenant à cette zone. Le véhicule est autorisé à sortir de la zone uniquement si l'entité qu'il transporte doit subir une opération dans une autre zone (i.e. si l'entité doit être livrée dans une autre zone). Dans ce cas, le véhicule change de zone et doit désormais servir les demandes émises par les points appartenant à cette nouvelle zone de service. D'autre part, si dans une zone il n'y a plus de demandes de transport attendant un véhicule, celui-ci va dans une zone de parking où il attend qu'une nouvelle requête soit émise par n'importe quel point du réseau, requête qu'il est autorisé à servir.

Pour le moment, nous allons comparer ces deux politiques de gestion de l'AMHS et analyser leur impact sur les performances du système de fabrication. Au chapitre 6, nous analyserons une méthodologie formelle pour la définition de zones dans la fab, pour ensuite présenter le couplage de cette démarche d'optimisation avec le modèle de simulation.

5.2.3 Plan d'expériences

L'hypothèse fondamentale d'un plan d'expériences consiste à dire que les valeurs des variables d'entrée du modèle influencent les variables de sortie. Ainsi, l'objectif est de construire des modèles permettant de prédire les valeurs des variables de sortie (e.g. le temps de cycle). Par exemple, considérons deux facteurs X_1 et X_2, et supposons qu'ils affectent la variable de sortie Y. Cette relation peut être approchée à l'aide du modèle de régression :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \beta_4 X_1^2 + \beta_5 X_2^2 + \epsilon \tag{5.1}$$

Où les coefficients β_j sont inconnus et doivent être estimés, et ϵ est l'erreur associée à l'inadéquation du modèle à estimer la véritable valeur de la variable de réponse Y. Les paramètres du modèle sont estimés à partir des répliques (scenarii) du modèle de simulation du système en donnant plusieurs valeurs pour les X_j, en enregistrant les valeurs solution

et ensuite en utilisant des modèles de régression pour leur estimation. Pour certains outputs de la simulation, il est raisonnable de supposer que la variation aléatoire de la valeur de l'indicateur de performance suit une loi de distribution normale. Lorsque nous nous intéressons à la valeur moyenne, nous pouvons appliquer le Théorème Central Limite et ainsi la distribution normale est une très bonne approximation pour les valeurs de ϵ [49, 6]. Selon le système étudié et les objectifs de l'étude, il peut y avoir autant de modèles de type (5.1) que de fonctions objectif (e.g. temps de cycle, utilisation des ressources). Ces modèles n'ont pas forcément les mêmes facteurs ni les mêmes termes.
Le lecteur intéressé par une description plus détaillée à propos des plans d'expériences pour la simulation et/ou de la méthodologie pour l'analyse des résultats peut se référer à [64, 6, 50].

L'objectif de l'analyse n'est pas forcément de présenter les modèles de régression type (5.1) pour chaque variable de sortie. Son intérêt concerne la formalisation des relations entre les variables d'entrée et de sortie de la simulation. Dans notre cas, notre intérêt est de regarder les effets que différents facteurs (variables d'entrée) pourraient avoir sur les mesures de performance de la fab (variables de sortie). Lors de notre analyse des résultats, nous allons donc nous concentrer sur les variations des indicateurs de performance du modèle dues à chacun des facteurs présentés ci-après au tableau 5.1.

Dans notre étude, les principaux facteurs qui peuvent avoir des effets sur les indicateurs de performance sont la règle de dispatching des véhicules, la politique de gestion par zones et la taille de la flotte. Les différents niveaux pour chacun des facteurs dans l'expérience sont présentés au tableau 5.1.

Tab. 5.1 – Facteurs et niveaux du plan d'expériences

Facteurs	Niveaux
Règle de sélection d'une tâche	FCFS, STD, CL-FCFS
Règle de sélection d'un véhicule	LIV
Nombre de zones dans la fab	1, 3
Nombre de véhicules	18, 24, 30, 36

Pour le moment, l'intérêt de cette étude de simulation est de mesurer l'impact sur les performances de la fabrication de la mise en œuvre des politiques de gestion du transport énnoncées. Nous avons donc conçu un plan d'expérience à facteurs multiples sans répétitions [81], avec un total de 24 ($3 \times 1 \times 2 \times 4$) scénarii. Pour chaque scénario, nous avons simulé la production de 4 mois. Les deux premiers mois ont été considérés comme période transitoire et les statistiques n'ont pas été prises en compte pour l'analyse des résultats afin d'éliminer des données biaisées dues à l'initialisation du modèle.

5.2.4 Résultats

Cette section est dédiée à la présentation des résultats des simulations. Étant donnés les importants investissements en machines et équipements lors de la construction des unités de production microélectroniques, les mesures de performance retenues dans cette étude

se basent principalement sur le temps de production. Remarquons que, afin de préserver la confidentialité de certaines informations industrielles, nous ne pouvons pas présenter la valeur absolue de certains indicateurs. Ceux-ci seront donc présentés sous forme de pourcentage. En particulier, nous nous sommes intéressés à analyser les indicateurs suivants :

- Le temps de cycle : C'est la somme des temps effectifs de production (temps total de process), temps d'attente devant les machines, temps effectif de transport, temps d'attente pour un véhicules et temps d'attente dans les stockers. Nous nous sommes intéressés à la valeur moyenne et son écart-type.
- Le taux d'utilisation maximal des machines : Le taux d'utilisation d'une machine est le rapport entre le temps total dédié à l'exécution des opérations sur les wafers et le temps total où la machine est disponible (hors temps dédié à la maintenance préventive ou réactive). Nous nous intéressons à la valeur maximale parmi toutes les machines dans la fab. Cet indicateur représente donc le niveau d'utilisation de la machine goulot qui détermine ainsi le rythme de production (débit de la fab) lorsque l'AMHS n'est pas une ressource bloquante pour la fabrication.
- Le taux d'utilisation des véhicules : C'est le pourcentage du temps simulé au cours duquel les véhicules réalisent un déplacement complet (à vide pour chercher une entité et en charge pour délivrer l'entité).
- Le temps de livraison : Ce temps correspond à la durée moyenne que demande un chariot pour effectuer un déplacement complet entre deux points quelconques du réseau. C'est-à-dire, pour n'importe quelle demande de transfert, c'est la somme du temps de déplacement à vide (pour chercher l'entité) plus le temps de déplacement en charge (pour délivrer cette entité).

Les résultats des simulations sont présentés dans le tableau 5.2. Pour des raisons de confidentialité, nous ne pouvons pas présenter les valeurs absolues pour le temps de cycle, nous donnons uniquement la valeur de son écart-type et remplaçons la moyenne par le X-factor. Comme expliqué au chapitre 2, le X-factor est défini par la relation entre le temps de cycle et le temps total de process ($\lfloor(\textit{temps de cycle})/(\textit{temps total de process})\rfloor$) et donne une bonne représentation du pourcentage du temps de cycle dû aux activités de non-production. L'objectif est donc d'avoir un X-factor aussi proche que possible de 1. Le tableau 5.2 présente également les valeurs du temps d'attente d'un véhicule en pourcentage du temps total de non-production. De plus, les figures 5.2, 5.3 et 5.4 reprennent les résultats obtenus pour le temps de cycle (sans les valeurs chiffrées) et le taux d'utilisation des véhicules pour chacun des scenarii considérés. Ces graphiques sont présentés à la même échelle afin de pouvoir comparer les valeurs du temps de cycle.

Les résultats de notre étude concernent fondamentalement des aspects peu évidents lorsque le modèle de simulation est une représentation trop simplifiée du système réel étudié, comme il est courant de le voir dans la littérature. Puisque notre modèle est une représentation assez détaillée de l'unité de fabrication, il nous a permis de constater que, sous certaines conditions, les bonnes performances obtenues à partir de l'implémentation d'une règle de dispatching ne sont plus valables lorsque certains paramètres du modèle changent (i.e. la taille de la flotte). De même, un résultat intuitif mais très important, concerne l'impact que peuvent avoir les règles de dispatching sur le nombre de véhicules nécessaires dans le système. Comme constaté dans d'autres études dans la littérature (e.g. [16]), il serait possible de réduire l'investissement en nombre de chariots si une bonne règle de dispatching était implémentée. En particulier, pour notre fab gérée avec une seule zone (politique "1-zone"), nous observons les très mauvaises performances de la règle FCFS, en

TAB. 5.2 – Résultats des simulations

Règle FCFS								
	1-zone				3-zone			
Nbr. véhicules	18	24	30	36	18	24	30	36
X-factor	15,89	8,66	3,49	1,34	10,46	5,10	1,94	1,33
Ecart-type du temps de cycle	13,33	8,99	3,47	0,04	10,90	5,44	0,97	0,04
Temps d'attente d'un véhicule (% du temps de non-production)	99%	93%	88%	50%	98%	54%	79%	49%
Utilisation maximale des machines	68%	74%	87%	93%	71%	78%	91%	93%
Utilisation de véhicules	100%	100%	100%	93%	100%	100%	100%	86%
Règle STD								
	1-zone				3-zone			
Nbr. véhicules	18	24	30	36	18	24	30	36
X-factor	1,56	1,31	1,30	1,30	1,81	1,34	1,30	1,29
Ecart-type du temps de cycle	0,17	0,04	0,04	0,04	0,60	0,05	0,04	0,04
Temps d'attente d'un véhicule (% du temps de non-production)	64%	44%	43%	43%	67%	50%	44%	42%
Utilisation maximale des machines	93%	93%	93%	93%	92%	93%	93%	93%
Utilisation de véhicules	100%	99%	93%	83%	100%	97%	88%	78%
Règle CL-FCFS								
	1-zone				3-zone			
Nbr. véhicules	18	24	30	36	18	24	30	36
X-factor	1,94	1,48	1,36	1,32	2,68	1,45	1,34	1,31
Ecart-type du temps de cycle	0,16	0,05	0,04	0,04	1,89	0,05	0,04	0,04
Temps d'attente d'un véhicule (% du temps de non-production)	79%	65%	53%	47%	72%	63%	51%	46%
Utilisation maximale des machines	93%	93%	93%	93%	87%	93%	93%	93%
Utilisation de véhicules	100%	100%	100%	90%	100%	100%	94%	83%

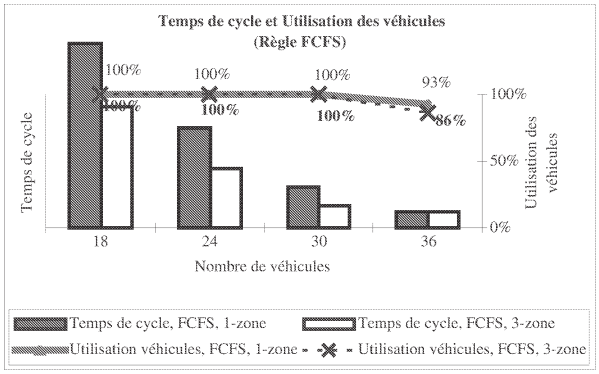


FIG. 5.2 – Temps de cycle et utilisation des véhicules, règle FCFS

particulier lorsqu'il y a peu de véhicules dans le système. Les performances de la règle FCFS s'améliorent en augmentant le nombre de chariots. Puisque cette règle est utilisée par les fournisseurs dans les systèmes de transport automatisés qu'ils installent et dans la plupart des travaux dans la littérature [91], cela pourrait justifier le fait de trouver un

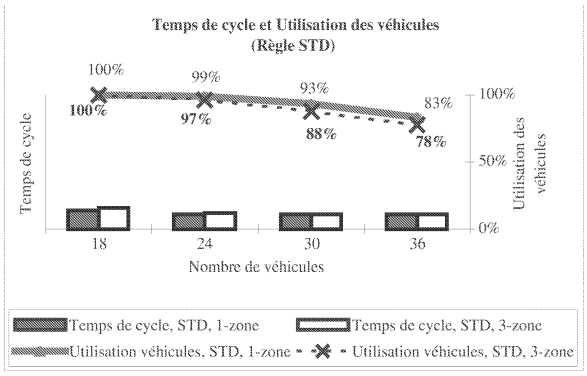


FIG. 5.3 – Temps de cycle et utilisation des véhicules, règle STD

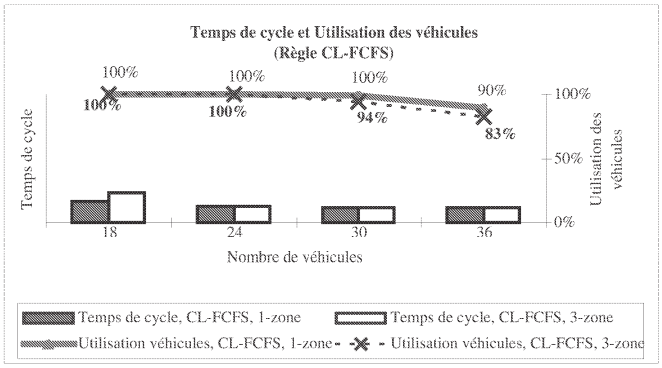


FIG. 5.4 – Temps de cycle et utilisation des véhicules, règle CL-FCFS

nombre important de véhicules dans les systèmes réels. Observons qu'en changeant la règle de dispatching, nous pouvons améliorer les performances de l'AMHS. En effet, avec 24 véhicules et la règle STD, nous pouvons obtenir le même résultat pour le temps de cycle (en termes de moyenne, écart-type et X-factor), la même utilisation maximale des machines et un meilleur taux d'utilisation des véhicules qu'avec 36 chariots et la règle FCFS. L'implémentation de la règle STD permet donc une réduction des investissements en moyens de transport. La mise en place de la règle hybride CL-FCFS permet d'améliorer certains indicateurs, mais elle n'est pas aussi performante que STD. Ainsi, nous pouvons voir qu'avec la politique "1-zone", la règle de dispatching STD donne les meilleurs résultats pour tous les indicateurs de performance considérés, et ce quelle que soit la taille de la flotte de véhicules.

Si nous regardons maintenant l'impact de la mise en place d'une stratégie de gestion par zones du système de transport sur les performances de la fab, nos résultats montrent qu'une stratégie de division du réseau par zones permet de réduire notablement le nombre de véhicules, pour une règle de dispatching donnée. En particulier, si la fab est divisée en zones, nous avons pu observer une diminution jusqu'à 69% du temps d'attente d'un chariot pour la règle FCFS avec 24 véhicules, de 5% avec la même règle et 36 véhicules. Pour ce qui est de la règle CL-FCFS, les niveaux d'amélioration vont de 5% à 11% selon le nombre de chariots. Avec la règle STD, nous pouvons arriver à une amélioration de 3% avec une flotte de 36 véhicules. Nous observons, cependant, une légère augmentation de 0,5% du temps

de cycle avec la mise en place de zones, le dispatching avec STD et 30 véhicules, mais le taux d'utilisation des véhicules est réduit de 6%. Ainsi, ce phénomène d'amélioration de l'utilisation des chariots se reproduit pour les autres scenarii. Ces améliorations du temps de cycle et du taux d'utilisation des véhicules sont beaucoup plus marquées lorsqu'il y a peu de véhicules dans le système. Ce gain de temps de non-production est ensuite ajouté au temps d'attente devant les machines ou en stockage dans les étagères, ce qui implique une meilleure utilisation des machines et nécessite donc la mise en place de règles plus sophistiquées pour l'ordonnancement de la production.

En ce qui concerne le temps effectif de transport (déplacement en charge pour délivrer une entité), la figure 5.5 présente un graphique qui compare les valeurs obtenues en fonction de la taille de la flotte des véhicules. Nous pouvons ainsi constater que le temps de transport ne varie pas, ce qui veut dire que le chemin le plus court est donc un bon choix pour le routage des véhicules. De plus, cela nous permet de constater qu'il n'y a pas de problèmes de congestions à l'intérieur du réseau.

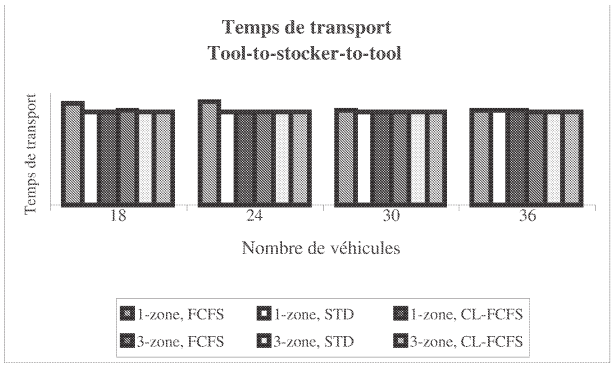


FIG. 5.5 – Temps de transport

Le tableau 5.3 présente les moyennes du temps de livraison pour chacune des règles de dispatching testées et les deux stratégies de gestion du transport par zones. Comme défini précédemment, ce temps correspond à la somme du temps de déplacement à vide pour chercher une entité plus le temps de déplacement en charge (pour délivrer l'entité). Le premier constat est que pour toutes les règles de dispatching, il est possible de diminuer le temps moyen d'une livraison lorsque la fab est divisée en zones. Cette amélioration en pourcentage est plus marquée pour la règle FCFS, montrant ainsi la pertinence de la mise en place d'une gestion par zones du système. De cette façon, cet indicateur peut être utilisé par les fabricants de semi-conducteurs pour négocier leurs contrats avec les fournisseurs des AMHS qui devront donc assurer ce taux de service.

TAB. 5.3 – Temps moyen de livraison

	1-zone	3-zone	Ecart
Règle FCFS	2,81	2,52	10%
Règle STD	2,21	2,10	5%
Règle CL-FCFS	2,33	2,22	5%

5.3 Conclusion

Ce chapitre a été consacré à la présentation de l'unité de fabrication de wafers étudiée dans cette recherche et à une étude de simulation ayant pour objectif principal l'analyse de certaines règles de dispatching des véhicules et son impact sur la performance globale de la fab. Notre étude expérimentale a montré les effets positifs d'une politique de gestion du transport par zones.

Nous avons vu au chapitre 3 que la règle FCFS est très souvent utilisée dans les travaux analysant les systèmes de transport dans les fabs. Dans notre étude, cependant, nous avons vu que cette règle a de très mauvaises performances sous certaines conditions. Afin d'améliorer les indicateurs de la fab lors de l'utilisation de FCFS pour le dispatching des véhicules, deux possibilités existent : soit on augmente le nombre de véhicules dans le système, soit on met en place une gestion du transport par zones. Dans le premier cas, l'investissement est important, alors que dans le second, il est possible d'avoir un temps de cycle de fabrication et un taux d'utilisation des véhicules et des machines tout à fait corrects. Nous avons testé d'autres règles de dispatching de véhicules (STD et une règle hybride CL-FCFS). Les résultats obtenus sont très intéressants, nous permettant d'obtenir des meilleures performances tout en diminuant le nombre de véhicules dans le système.
Nous avons aussi testé l'impact d'une politique de gestion de l'AMHS par zones. Avec cette politique, les zones ont été définies selon la localisation géographique des points dans le réseau, lesquelles nous ont permit de traiter les demandes de transport. Nous avons observé une amélioration des principaux indicateurs, permettant une réduction du taux d'utilisation des chariots pour un temps de cycle équivalent, le nombre de véhicules pouvant être également diminué.

Les résultats obtenus en simulation sont très encourageants pour proposer une démarche formelle pour la définition de zones dans la fab. A partir du chapitre suivant, dans la seconde partie de cette thèse, nous allons proposer une démarche d'optimisation pour la définition des zones et l'affectation de véhicules à chacune des zones. Cette dimension du problème constitue le niveau tactique. Ensuite, au niveau opérationnel, la solution de l'optimisation sera couplée avec notre modèle de simulation de la fab dans le but d'étudier son comportement dynamique et les performances opérationnelles.

Deuxième partie

Optimisation de la gestion du système de transport automatisé

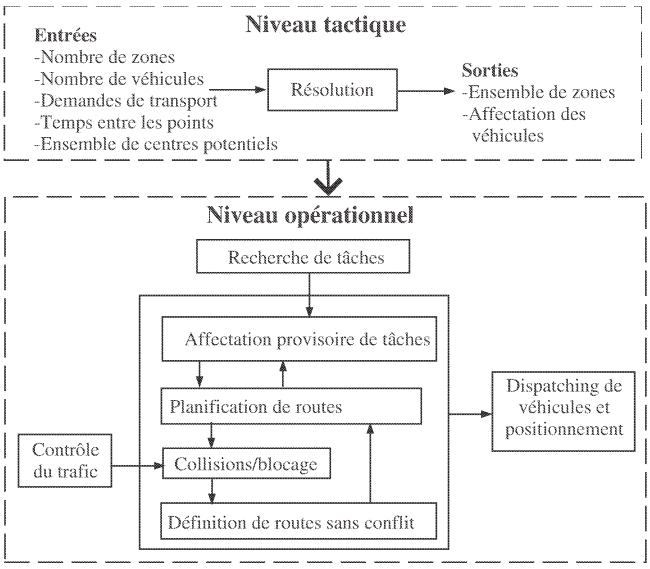
Préambule

La première partie de cette thèse était dédiée à la présentation du contexte industriel et économique, à la compréhension de la problématique et à la modélisation et l'analyse du cas industriel. Nous avons ainsi pu voir les caractéristiques du processus de fabrication de semi-conducteurs qui en font un véritable challenge pour la recherche en génie industriel. Le centre d'intérêt de notre travail étant l'étude du système de transport automatisé de l'unité de fabrication, nous avons réalisé, à la fin de la première partie de cette thèse (chapitre 5), une étude par simulation de deux stratégies de gestion du système de transport et de différentes règles pour le dispatching des véhicules. Nous avons ainsi pu analyser le comportement de la fab sous certaines conditions, identifier les paramètres critiques, et tester, en particulier, la pertinence d'une gestion par zones du réseau de transport. Nous avons ainsi pu voir que la performance du système de transport peut être améliorée soit en augmentant le nombre de véhicules, soit en découpant le réseau en zones de gestion. Puisque l'augmentation du nombre de chariots nécessite des investissements importants, il est donc préférable de trouver des solutions moins coûteuses, ce qui est le cas du découpage en zones. De plus, avec une flotte de véhicules donnée, le découpage en zones permet de mieux utiliser les ressources de transport (i.e. diminuer le taux d'utilisation des véhicules), ce qui est très favorable pour leur gestion pendant les périodes de pannes ou de maintenance des chariots.

Dans la seconde partie de cette thèse, nous allons proposer une démarche d'optimisation pour la définition des zones, l'affectation de véhicules à chacune des zones et la gestion opérationnelle des véhicules dans le réseau. La figure ci-après présente les liens entre ces aspects de la problématique de gestion du système de transport. Nous pouvons identifier deux niveaux : un niveau tactique et un niveau opérationnel.

La dimension tactique du problème est analysée au chapitre 6. L'objectif est alors de découper le réseau de transport en zones de gestion et d'affecter une flotte de véhicules à chacune des zones. A ce niveau, les informations sur la demande et le comportement des véhicules sont prises en compte de manière agrégée. Nous appréhendons le problème à l'aide de la formulation de programmes linéaires en nombres entiers. A terme, l'objectif est de prendre en compte cette dimension du problème dans une démarche itérative d'aide à la décision.

Le chapitre 7 est dédié au niveau opérationnel. L'objectif est alors de définir des politiques pertinentes pour la gestion opérationnelle des véhicules dans la fab. La solution de l'optimisation du niveau tactique sera donc intégrée au modèle de simulation de la fab afin d'étudier son comportement dynamique et les performances opérationnelles. D'autres aspects, comme par exemple le routage des chariots ou la gestion du trafic, doivent aussi être



pris en compte dans la simulation. Les liens entre les deux dimensions sont aussi analysés dans ce chapitre. En particulier, nous nous intéressons à l'impact du dimensionnement des paramètres comme le nombre de zones et le nombre de véhicules.

Chapitre 6

Définition tactique des zones

Nous avons vu dans le chapitre précédent que la performance du système de transport peut être améliorée soit en augmentant le nombre de véhicules, soit en découpant le réseau en zones de gestion. Puisque l'augmentation du nombre de chariots nécessite des investissements importants, il est donc préférable de trouver des solutions moins coûteuses, ce qui est le cas du découpage en zones. De plus, avec une flotte de véhicules donnée, le découpage en zones permet de mieux utiliser les chariots (i.e. diminuer leur taux d'utilisation), ce qui est très favorable pour la gestion lors des pannes. Dans ce chapitre, nous allons analyser plus formellement le problème du découpage du système en zones. A l'aide de modèles mathématiques, le problème est traité du point de vue tactique, c'est-à-dire, préalablement à la mise en place opérationnelle de politiques de gestion dynamique de véhicules. Nous proposons également une résolution efficace à travers une méthode heuristique.

6.1 Problème tactique de positionnement des véhicules

Comme expliqué précédemment dans cette thèse, l'objectif du positionnement tactique de véhicules est de trouver un emplacement "idéal" des véhicules dans le réseau de manière à optimiser un certain critère (e.g. le temps maximal d'attente d'un véhicule par une entité). L'idée est de découper notre système de production en plusieurs zones (espace logique qui regroupe un ensemble de machines) et d'affecter à chaque zone une flotte de véhicules qui assurera le transport de produits entre machines (points du réseau). Un exemple de découpage du réseau de transport est illustré dans la figure 6.1. Ici, le réseau a été découpé en 5 zones, et 24 véhicules ont été affectés (3 véhicules pour chacune des zones 1, 2 et 3, 6 véhicules pour la zone 4 et enfin 9 véhicules pour la zone 5). Les zones sont définies par leurs centres, les points qu'elles contiennent et les véhicules qui leur sont affectés. Les centres sont des points du système. Dans un souci de simplification, on considère, au niveau tactique, que les véhicules appartenant à une zone stationnent au centre de la zone. Comme expliqué dans notre revue de la littérature au chapitre 3, le problème de positionnment de véhicules dans les systèmes automatisés de production a traditionnellement consisté à trouver un emplacement des chariots lorsqu'ils se libèrent et qu'il n'y a pas de requête en attente. Notre objectif dans cette thèse est de distribuer les véhicules dans le réseau de telle façon que le système soit couvert et les requêtes puissent être satisfaites dans les meilleurs délais. Il s'agit donc de toujours avoir des véhicules près des endroits où il y aura une demande de transport. A notre connaissance, cette approche du problème de positionnement de véhicules (i.e. découpage du réseau en zones de gestion et affectation

d'une flotte à chaque zone) n'a pas été étudiée auparavant.

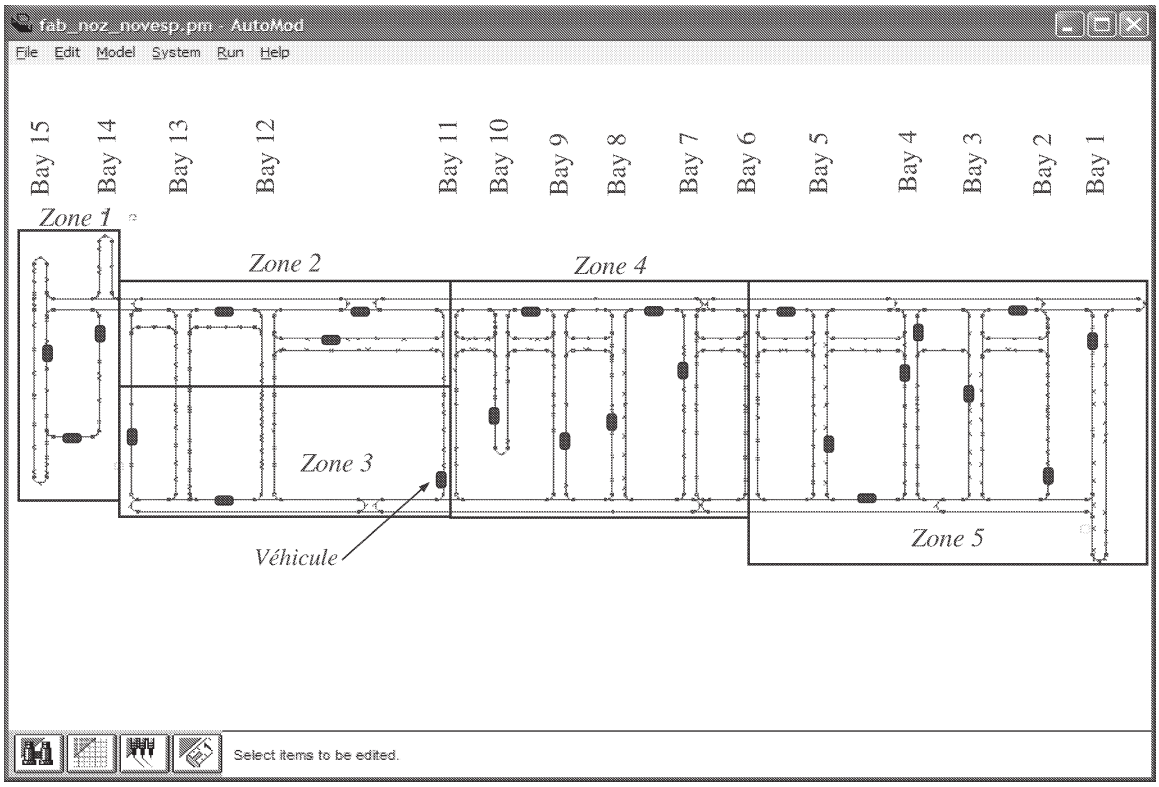


FIG. 6.1 – Exemple de découpage du réseau de transport en zones

6.2 Formulations mathématiques

6.2.1 Considérations pour la modélisation et hypothèses

Le développement des modèles mathématiques est basé sur les hypothèses suivantes :

1. La disposition des machines et les flux de transport sont connus, les distances entre les différents points, l'emplacement des machines et les sens de parcours sont fixés.
2. Les flux de produits en termes de charge par période sont connus, ils sont calculés en fonction de la gamme opératoire (l'ordre des opérations) des produits à fabriquer et leurs demandes sur l'horizon de fabrication.
3. Chaque fois qu'un véhicule visite une machine, il existe toujours du produit à transférer. Autrement dit, les déplacements à vide ne sont effectués que lorsqu'un véhicule doit satisfaire une requête.
4. Les véhicules sont supposés identiques (e.g. ils ont la même vitesse, accélération) et une capacité de charge unitaire (i.e. ils ne peuvent transporter qu'une entité à la fois).

5. Pour le problème de gestion du trafic, le contrôle aux points d'intersection du réseau unidirectionnel est suffisant pour éviter les collisions.
6. On suppose que les transporteurs et les machines sont fiables.
7. Le nombre de produits à fabriquer pendant une période donnée est connu et constant sur tout l'horizon de production. Cette hypothèse est réaliste quand la politique de production est définie à des niveaux hiérarchiques plus élevés. Cela nous permet aussi de formuler la fonction objectif pour minimiser le temps maximal d'attente d'un véhicule par une entité.

Ces hypothèses seront discutées dans la section 6.3, et on verra que certaines peuvent aisément être relachées.

D'autre part, comme nous avons vu au chapitre 5, le réseau de transport considéré dans cette étude est uni-directionnel, et les véhicules ont une capacité de charge unitaire. Ces caractéristiques ne constituent pas de contraintes à l'applicabilité de nos modèles pour la résolution du niveau tactique de gestion. Nous pouvons appliquer les modèles proposés ci-après sur des réseaux bi-directionnels ou sur des systèmes avec des véhicules à capacité multiple ou bi-directionnels. Ces modifications ne sont problématiques qu'au niveau opérationnel pour la mise en place de règles de dispatching et pour la gestion du trafic. En effet, si la capacité de charge par véhicule est supérieure à une unité, la gestion opérationnelle devient beaucoup plus complexe [112, 9, 106] car nous allons devoir considérer le fait qu'un véhicule peut, pendant son parcours de livraison, aller chercher une autre entité si sa capacité totale de charge le lui permet. Le problème est plus complexe si certains véhicules peuvent transporter plus d'une entité à la fois alors que d'autres non. Si nous considérons le réseau ou les véhicules comme étant bi-directionnels, alors la gestion du trafic deviendrait, elle aussi, plus complexe [79] car il faudrait mettre en place notamment des règles de gestion visant à éviter les collisions frontales entre les véhicules.

6.2.2 Programmes linéaires en nombre entiers (PLNE)

6.2.2.1 Minimisation du temps maximal de couverture

Dans un premier temps le problème d'optimisation consiste à minimiser le temps de couverture maximal, c'est-à-dire à minimiser le temps que nécessite un véhicule pour effectuer un déplacement entre le point nommé "centre" de la zone et le point le plus éloigné. Le choix de ce critère s'explique par le fait que la performance du système est fortement liée au temps nécessaire à un véhicule pour répondre à une demande. Le point "centre" de la zone est supposé être le point sur lequel les véhicules sont stationnés lorsqu'il n'y pas de travail à effectuer dans la zone (i.e. il n'y a pas de demandes de transport en attente). Comme nous le verrons plus tard, la notion de "centre" est utilisée dans le même sens que dans les problèmes classiques de localisation d'installations.

Pour formuler le problème nous précisons les notations suivantes :

Paramètres

– P : nombre maximum de zones,

- V : ensemble des véhicules,
- K : ensemble des points de chargement/déchargement du réseau (machines),
- J : ensemble des points du réseau qui peuvent être pris comme centre de zones,
- t_{ik} : temps de parcours entre les points i et k du réseau,
- D_k : demande moyenne du point k définie comme le temps moyen (en secondes) par heure nécessaire pour réaliser tous les déplacements au départ du point k,
- VC : capacité de véhicules calculée comme le temps de disponibilité (en secondes) d'un véhicule par heure.

Variables

- T : temps de couverture,
- $X_j = 1$ si le point j est choisi comme centre d'une zone, 0 sinon,
- $Y_{jk} = 1$ si le point k appartient à la zone dont j est le centre, 0 sinon,
- $Z_{vj} = 1$ si le véhicule v est assigné à la zone dont j est le centre, 0 sinon.

La fonction objectif est exprimée comme : Minimiser $\underset{k \in K}{\text{Maximum}} \sum_{j \in J} t_{jk} Y_{jk}$. Nous pouvons modéliser notre problème comme un programme linéaire en nombre entiers :

Modèle VP1 :

$$\text{Minimiser } T \tag{6.1}$$

Sous :

$$T \geq \sum_{j \in J} t_{jk} Y_{jk} \qquad \forall k \in K \tag{6.2}$$

$$\sum_{j \in J} X_j \leq P \tag{6.3}$$

$$\sum_{j \in J} Y_{jk} = 1 \qquad \forall k \in K \tag{6.4}$$

$$Y_{jk} \leq X_j \qquad \forall j \in J, k \in K \tag{6.5}$$

$$\sum_{j \in J} Z_{vj} \leq 1 \qquad \forall v \in V \tag{6.6}$$

$$\sum_{k \in K} D_k Y_{jk} \leq VC \sum_{v \in V} Z_{vj} \qquad \forall j \in J \tag{6.7}$$

$$X_j \in \{0, 1\} \qquad \forall j \in J \tag{6.8}$$

$$Y_{jk} \in \{0, 1\} \qquad \forall j \in J, k \in K \tag{6.9}$$

$$Z_{vj} \in \{0, 1\} \qquad \forall v \in V, j \in J \tag{6.10}$$

Comme souligné plus haut, l'objectif est de diviser le réseau AMHS en zones et d'assigner à chaque zone une flotte de véhicules. La fonction objectif (6.1) et l'ensemble des contraintes (6.2) cherchent à minimiser le temps maximum pour couvrir une demande. Les

contraintes (6.3) assurent que le nombre maximal de zones n'est pas dépassé. Les ensembles de contraintes (6.4) et (6.5) établissent respectivement, que chaque point du réseau appartient à une zone et une seule, et qu'un point n'est couvert par une zone que si seulement cette zone est créée. L'ensemble des contraintes (6.6) garantit qu'un véhicule est assigné au plus à une zone. Les contraintes (6.7) s'assurent que la capacité de chaque zone (relatif au nombre de véhicules dans la zone) est satisfaite. Les contraintes (6.8), (6.9) et (6.10) conditionnent le fait que les variables de décision soient binaires.

Cette formulation du problème est relativement semblable au problème du P-center, qui est un problème classique de localisation d'installations [22]. Notre formulation, néanmoins, a les caractéristiqeus suivantes qui la différencient des modèles classiques de localisation d'installations :

1. Découpage et distribution des ressources. Dans le problème P-center, l'objectif est de trouver la localisation de P installations qui minimise la distance (ou le temps) maximale de couverture. Les P installations sont en fait les ressources qu'il faut distribuer pour satisfaire les demandes des clients. Dans notre cas, nous voulons créer P zones (i.e. diviser le réseau de transport en P zones). Pour nous, la demande des points du réseau (les clients dans le P-center) est satisfaite par une flotte de V véhicules. Ainsi, dans notre cas, nos ressources sont les véhicules qui doivent être distribués dans les zones de couverture créées.
2. Contraintes de capacité. Notre formulation est une extension du problème du P-center qui prend en compte des contraintes de capacité. Nous nous rapprochons donc du *Capacitated Facility Location Problem* (CFLP) mais nous n'avons pas de coûts liés à l'ouverture d'une installation (i.e. choix d'un centre de zone). Notre objectif est celui du P-center.
3. Capacité globale du réseau. La version standarde du CFLP considère les capacités des installations de façon séparée, individuellement pour chaque installation potientielle. Dans notre formulation, nous nous intéressons à une distribution globale de la capacité dans les zones. En effet, notre seule contrainte sur la capacité est la taille totale de la flotte de véhicules. Ainsi, la création des zones est effectuée selon la distribution de la demande et l'affectation des véhicules (distribution de la capacité) est réalisée selon les besoins spécifiques pour chaque zone tout en respectant la contrainte globale (nombre limité de véhicules dans le réseau).
4. Capacité discrète. Dans certains cas, la formulation du CFLP peut conduire à la résolution d'un problème avec capacité continue (e.g. selon le produit à délivrer aux clients du réseau). Dans notre cas, nous avons toujours la certitude d'avoir des capacités discrètes car nos ressources sont en fait les véhicules disponibles qu'il faut repartir dans les différentes zones créées.
5. Enfin, comme nous le verrons dans la suite de ce chapitre, puisque nous nous intéressons à la distribution de la capacité globale dans les zones créées, nous ajoutons des critères à la fonction objectif afin d'équilibrer les demandes et d'équilibrer les ressources (véhicules) dans tout le réseau.

Pour garantir une solution réalisable, la capacité totale des véhicules dans le système doit être plus grande que la demande totale, ce qui signifie que l'équation (6.11) doit être satisfaite. Elle permet aussi d'obtenir la taille minimale de la flotte de véhicules.

$$\sum_{k \in K} D_k \leq VC \times |V| \tag{6.11}$$

Ajout de coupes

Afin d'améliorer le temps de calcul des modèles avec l'utilisation d'un logiciel standard, sans pour autant modifier la valeur de la fonction objectif (i.e. le temps de couverture maximal), des coupes peuvent être ajoutées. Ces coupes sont représentées par les contraintes (6.12), (6.13), (6.14) et (6.15).
Dans le but de garantir l'obtention de solutions plus réalistes, les contraintes (6.12) assurent qu'un véhicule ne soit pas affecté à une zone si aucun point n'est assigné à cette zone. Les contraintes (6.13) assurent qu'une zone ne peut être ouverte si aucun point n'est assigné à cette zone. Les contraintes (6.14) assurent que si une zone est ouverte, alors au moins un véhicule doit être affecté à cette zone.

$$Z_{vj} \leq \sum_{k \in K} Y_{jk} \qquad \forall v \in V, j \in J \tag{6.12}$$

$$X_j \leq \sum_{k \in K} Y_{jk} \qquad \forall j \in J \tag{6.13}$$

$$X_j \leq \sum_{v \in V} Z_{vj} \qquad \forall j \in J \tag{6.14}$$

Lors de l'affectation des véhicules aux zones, il existe des solutions symétriques. Pour expliquer la nature du problème de symétrie, supposons que nous voulions créer 2 zones et que nous disposions de 3 véhicules (v_1, v_2 et v_3). Les zones ont été créées de telle façon que la demande de la zone 1 impose l'affectation de 2 véhicules, alors que la demande de la zone 2 impose l'affectation de 1 seul véhicule. Sous ces conditions, nous pouvons affecter les véhicules comme illustré au tableau 6.1.

TAB. 6.1 – Exemple du problème de symétrie

	Zone	Véhicules affectés
1ère possibilité :	Zone 1	v_1, v_2
	Zone 2	v_3
2ème possibilité :	Zone 1	v_1, v_3
	Zone 2	v_2
3ème possibilité :	Zone 1	v_2, v_3
	Zone 2	v_1

Puisque nous considérons que tous les véhicules sont identiques, il n'est pas nécessaire d'évaluer toutes ces possibilités. Une fois les véhicules v_1 et v_2 ont été affectés, nous pouvons continuer immédiatement à évaluer l'affectation des véhicules restants aux autres zones. Formellement, le problème de symétrie est abordé dans notre modèle par les contraintes (6.15) présentées ci-dessous.

$$Z_{vj} \leq \sum_{k=1}^{j} Z_{v-1,k} \qquad \forall v = 2, ...|V|, j \in J \tag{6.15}$$

L'impact de ces coupes sur le temps de résolution du modèle sera analysé par la suite.

Résultats expérimentaux

Les programmes linéaires en nombre entiers ont été implémentés sur un PC Pentium® 4 (3,4 GHz) en utilisant le solver Xpress-MP® [21]. Nous avons réalisé une étude expérimentale avec les données fournies par notre partenaire industriel. Les valeurs des temps de transport entre les points du réseau sont obtenus à partir d'une matrice de dimension $|K| \times |K|$, où $|K| > 200$ représente le nombre de points de chargement/déchargement (i.e. nombre de machines dans la fab). L'ensemble J de centres potentiels de zone peut être identique à K. Cependant, à cause de contraintes informatiques (i.e. espace mémoire), nous avons défini $|J| = 30$ points préalablement à la résolution des programmes linéaires. Ces points ont été choisis en utilisant la méthode gloutonne proposée dans la section 6.4.1. L'objectif de l'utilisation de cette méthode était de choisir, parmi tous les points du réseau, un ensemble d'exactement 30 points qui donnent la configuration de 30 zones permettant de minimiser le temps maximal de couverture. Nous avons ensuite gardé les 30 points et les avons utilisés dans la résolution des PLNE. Remarquons que la solution optimale fournie par le modèle mathématique peut se voir affectée par le choix de ces 30 points, comme présenté au tableau 6.2 ci-après [89]. Trois ensembles différents ont été définis et les valeurs solution du temps maximal de couverture ont été enregistrées. Dans l'étude expériementale qui va suivre, nous avons gardé l'ensemble J.

TAB. 6.2 – Sensibilité du PLNE par rapport aux points dans J

	Nombre de zones		
Ensemble	5	10	15
J	73	50	41
J'	73	51	41
J''	63	63	63

En utilisant l'équation (6.11), nous pouvons calculer le nombre minimum de véhicules nécessaires. A partir des données fournies par l'entreprise, nous avons $|V| > 18$, ce résultat est un calcul statique. En effet, au chapitre 5, nous avons utilisé un modèle de simulation pour analyser le comportement dynamique de la fab, et nous avons pu constater que, afin de satisfaire dynamiquement toutes les demandes pendant l'horizon d'étude, une flotte supérieure à 18 véhicules était nécessaire (en fait, pour la plupart des politiques testées pour la gestion opérationnelle du système de transport (règle de dispatching des véhicules), nous avons eu besoin d'environ 24 véhicules). Nous avons donc testé notre modèle de PLNE avec une flotte de 24 et 30 véhicules.

Lors de la résolution avec les paramètres standards de ce logiciel, nous avons observé que l'exploration des solutions était réalisée en évaluant d'abord l'appartenance des points du réseau aux zones créées (variables Y_{jk}). Puisque l'objectif du modèle est de créer des zones en minimisant le temps de couverture maximal, il est préférable de décider d'abord de la sélection d'un point comme centre de zone (variable X_j) et ensuite de choisir la configuration des zones (valeurs des variables Y_{jk}) et enfin d'affecter la flotte de véhicules (variables Z_{vj}). Nous avons donc forcé le logiciel à effectuer cette évaluation en premier lieu à partir de la variable X_j et ensuite à partir de Y_{jk}. Les résultats obtenus pour les temps de calcul sont présentés au tableau 6.3. Avec cette modification des paramètres standards du logiciel, nous pouvons observer une amélioration rélative du temps de calcul d'entre 85% et 97% (soit une diminution de 7 à 33 fois).

TAB. 6.3 – Temps de résolution du modèle **VP1** (en sec., sans coupes)

Nombre de véhicules	24			30		
Nombre de zones	5	10	15	5	10	15
Avec paramètres standards du logiciel	2759	4176	2728	1794	1601	1408
Avec X_j prioritaires	197	482	190	134	241	43

Toujours afin d'améliorer le temps de calcul, nous avons étudié l'impact des coupes proposées précédemment. Les résultats obtenus sont présentés dans le tableau 6.4. Avec les coupes (6.13) et (6.14) les gains sont conséquents puisqu'ils vont de 11% à 89% avec 24 véhicules et de 10% à 40% avec une flotte de 30 véhicules. Avec les coupes (6.12) et (6.15), nous espérions gagner aussi en rapidité, mais cela n'a malheureusement pas toujours été le cas. Un gain de 49% a été obtenu avec la coupe (6.12) pour une configuration avec 24 véhicules et 15 zones. D'autre part, bien qu'avec la coupe (6.15) nous n'ayons pas gagné en rapidité, son intérêt concerne l'ordre dans lequel sont affectés les véhicules et donc la présentation de la solution.

TAB. 6.4 – Impact des coupes sur les temps de calcul (en sec.)

Nombre de véhicules	24			30		
Nombre de zones	5	10	15	5	10	15
Sans coupes	197	482	190	134	241	43
Avec (6.12)	339	767	96	581	739	197
Avec (6.13)	175	270	103	203	190*	30
Avec (6.14)	174*	265*	20*	120*	200	26*
Avec (6.12) et (6.13)	440	804	279	552	880	265
Avec (6.13) et (6.14)	207	333	49	156	248	38
Avec (6.12), (6.13) et (6.14)	428	689	307	539	669	79
Avec (6.15)	544	4398	26110	437	7612	4014

*Meilleurs temps

Le tableau 6.5 ci-après présente les résultats des tests numériques.

Nous pouvons observer qu'avec cette formulation, nous avons intérêt à ouvrir le nombre P maximal de zones. Cela nous permet de diminuer le temps de couverture maximal. L'affectation des véhicules est effectuée ensuite selon la demande de chaque zone. Ainsi, malgré

Tab. 6.5 – Résultats du modèle **VP1**

Nombre de véhicules : 24			
Nombre de zones maximum	5	10	15
Nombre de zones créées	5	10	15
Temps maximal de couverture	73	50	41
Utilisation moyenne des véhicules	78,5%	75,1%	68,4%
Ecart-type de l'utilisation des véhicules	12,7%	19,8%	23,7%
Utilisation minimale des véhicules	67,8%	38,3%	29%
Utilisation maximale des véhicules	99,9%	97%	97,3%
Demande moyenne des zones	12526,8	6263,4	4175,6
Ecart-type de la demande des zones	7873,3	4031	2680,1
Nombre de véhicules : 30			
Nombre de zones maximum	5	10	15
Nombre de zones créées	5	10	15
Temps maximal de couverture	73	50	41
Utilisation moyenne des véhicules	76,9%	76,7%	77,7%
Ecart-type de l'utilisation des véhicules	14,1%	17,4%	17,1%
Utilisation minimale des véhicules	57,1%	51%	44,9%
Utilisation maximale des véhicules	92,1%	99,9%	98%
Demande moyenne des zones	12526,8	6263,4	4175,8
Ecart-type de la demande des zones	9371,7	2085,3	3269,2

le fait que le taux moyen d'utilisation des véhicules soit acceptable, nous remarquons que dans certains cas, nous avons des véhicules qui seront utilisés à presque 100% (e.g. à 99,9% pour le cas avec 5 zones et 24 véhicules) alors que dans d'autres zones les véhicules seront sous-utilisés (e.g. à 67,8% pour le cas avec 5 zones et 24 véhicules). Ces résultats se confirment même lorsque nous avons un nombre de véhicules important dans le système (rappelons que dans les simulations effectuées au chapitre 5, une flotte de 30 véhicules était plus que suffisante pour satisfaire dynamiquement les demandes dans le réseau). De plus, dans les expériences que nous avons réalisées, nous observons que l'écart entre les valeurs maximale et minimale du taux d'utilisation des véhicules augmente au fur et à mesure que nous augmentons le nombre P de zones à créer. En effet, pour le cas avec 24 véhicules, cet écart est de 32,1% pour $P = 5$ zones, 58,7% pour $P = 10$ zones et 68,3% pour $P = 15$ zones. Pour les expériences avec 30 véhicules, l'écart entre les valeurs maximale et minimale de ce taux est de 35% pour $P = 5$, 48,9% pour $P = 10$ et de 53,1% pour $P = 15$ zones. Ce phénomène peut aussi être observé à partir de l'écart-type présenté dans le tableau.

Ces résultats sont principalement dus au fait que l'unique objectif du modèle mathématique **VP1** est de trouver une configuration de zones et une affectation des véhicules qui minimise le temps maximal de couverture (i.e. le temps que nécessite un véhicule pour aller du centre de la zone vers le point le plus éloigné de la zone). Le modèle ne s'intéresse pas à l'affectation des véhicules ou à l'équilibrage de la charge de travail (i.e. équilibrage de leur taux d'utilisation) delors qu'il y en a assez pour satisfaire les demandes des zones créées. Si par hasard (ou à cause d'une décision stratégique : achat d'un nombre donné de véhicules) la taille de la flotte de véhicules ne permet pas de satisfaire les demandes de toutes les P zones, alors ce modèle est contraint de créer un nombre de zones inférieur à P qui garantisse le minimum temps de couverture maximale.

6.2.2.2 Prise en compte du taux d'utilisation des véhicules

Comme expliqué précédemment, l'objectif du modèle **VP1** est de minimiser le temps maximal de couverture. Nous avons donc intérêt à ouvrir toutes les P zones, et peu importe que la charge de véhicules (exprimée par leur taux d'utilisation) soit très élevée ou très petite. A titre d'exemple, observons dans le tableau 6.5 la solution obtenue pour 30 véhicules et 15 zones. Nous avons ouvert les 15 zones, et les véhicules ont été affectés de façon à satisfaire la demande de chacune des zones, mais leur taux d'utilisation maximal et minimal sont, respectivement, 98% et 44,9%. Nous pourrions mieux distribuer les véhicules dans les zones de manière à ne pas avoir de véhicules trop chargés ou sous-utilisés.

De plus, dans le modèle **VP1**, nous supposons que le temps de couverture est minimisé avec une disponibilité des véhicules de 100% du temps (e.g. pas de prise en compte des pannes). Nous pouvons fixer le taux d'utilisation maximale des véhicules et l'inclure dans les contraintes du modèle. Le taux d'utilisation est donc minimisé en réduisant successivement sa valeur et en résolvant le modèle jusqu'à ce que le temps de couverture augmente ou que le modèle devienne infaisable. Losque le taux d'utilisation de véhicules est pris en compte dans les contraintes, nous avons besoin d'un paramètre additionnel noté UR_{max} ($UR_{max} \in [0,1]$). Maintenant, les contraintes (6.7) peuvent être remplacées par les contraintes (6.16) ci-dessous :

$$\sum_{k \in K} D_k Y_{jk} \leq UR_{max}\ VC \sum_{v \in V} Z_{vj} \qquad \forall j \in J \tag{6.16}$$

Dans le modèle **VP2** proposé ci-dessous, UR_{max} est considéré comme une variable et nous devons donc prendre en compte le taux d'utilisation individuel des véhicules. Un ensemble de variables UR_{vj} est défini comme étant le taux d'utilisation du véhicule v assigné à la zone dont le centre est le point j ($UR_{vj} \in [0,1]$). La fonction objectif est maintenant exprimée comme une fonction à objectifs multiples, où CO_1 est le poids associé au premier critère (temps de couverture), et CO_2 est le poids associé au second critère (taux maximal d'utilisation des véhicules).

Modèle VP2 :

$$\text{Minimiser}(CO_1.T + CO_2.UR_{max}) \tag{6.17}$$

Sous :

$$T \geq \sum_{j \in J} t_{jk} Y_{jk} \qquad \forall k \in K \qquad (6.18)$$

$$\sum_{j \in J} X_j \leq P \qquad (6.19)$$

$$\sum_{j \in J} Y_{jk} = 1 \qquad \forall k \in K \qquad (6.20)$$

$$Y_{jk} \leq X_j \qquad \forall j \in J, k \in K \qquad (6.21)$$

$$\sum_{j \in J} Z_{vj} \leq 1 \qquad \forall v \in V \qquad (6.22)$$

$$\sum_{k \in K} D_k Y_{jk} \leq VC \sum_{v \in V} UR_{vj} \qquad \forall j \in J \qquad (6.23)$$

$$UR_{vj} \leq Z_{vj} \qquad \forall v \in V, j \in J \qquad (6.24)$$

$$UR_{max} \geq \sum_{j \in J} UR_{vj} \qquad \forall v \in V \qquad (6.25)$$

$$X_j \in \{0, 1\} \qquad \forall j \in J \qquad (6.26)$$

$$Y_{jk} \in \{0, 1\} \qquad \forall j \in J, k \in K \qquad (6.27)$$

$$Z_{vj} \in \{0, 1\} \qquad \forall v \in V, j \in J \qquad (6.28)$$

$$UR_{vj} \in [0, 1] \qquad \forall v \in V, j \in J \qquad (6.29)$$

Dans ce modèle, les contraintes (6.16) du modèle précédent sont représentées par les ensembles de contraintes (6.23), (6.24) et (6.25). Les contraintes (6.23) dépendent maintenant de UR_{vj}, au lieu des Z_{vj} (i.e. des taux d'utilisation réels des véhicules assignés aux zones). Les contraintes (6.24) assurent que le taux d'utilisation d'un véhicule dans une zone n'est positif que si le véhicule est affecté à la zone. L'utilisation maximale des véhicules est obtenue à partir des contraintes (6.25). Les autres contraintes du modèle sont identiques à celles définies précédemment.

Le modèle **VP2** étant un modèle multi-objectif, on pourrait, lors de sa résolution, s'intéresser à la recherche des ensembles de solutions Pareto-optimales. Pour ce faire, on peut faire varier les valeurs des coefficients CO_1 et CO_2 de la fonction objectif. Cependant, notre objectif n'est pas de trouver les ensembles de solutions Pareto-optimales. Notre intérêt est de formaliser le problème et de voir l'impact d'une solution tactique de positionnement des véhicules sur la performance opérationnelle (par simulation) du système. Nous avons donc fixé les valeurs des coefficients $CO_1 = 1$ et $CO_2 = 1$. Ce choix est aussi justifié par le fait que les valeurs de T et de UR_{max} sont bornées : $1 \leq T \leq \max t_{ik}$ et $0 \leq UR_{max} \leq 1$. Ainsi, T pouvant être au moins égal à 1 et UR_{max} ne pouvant être supérieur à 1, le poids de T (le temps de couverture maximal) dans la fonction objectif sera toujours plus important que le taux d'utilisation des véhicules.

Résultats expérimentaux

Pour l'étude expérimentale du modèle **VP2**, nous avons utilisé les mêmes conditions de test que pour le modèle **VP1**. La première analyse réalisée concerne l'impact des coupes sur le

temps de calcul. Les résultats sont présentés au tableau 6.6. Les gains que nous pensions obtenir en ajoutant les coupes ne sont conséquent que dans quelques-uns des tests effectués. Comme expliqué plus haut, **VP2** étant un modèle mathématique à objectifs multiples, les valeurs des coefficients CO_1 et CO_2 de la fonction économique peuvent être modifiés afin de trouver les ensembles de solution Pareto-optimales. A partir des explications déjà données, nous avons fixé les valeurs des coefficients $CO_1 = CO_2 = 1$. Nous avons diminué le temps de résolution seulement pour trois instances parmi tous les cas testés. Ces gains peuvent aller par exemple jusqu'à 31% pour le cas avec 30 véhicules et 15 zones en ajoutant la coupe (6.14). Pour certains cas où il n'y a pas eu d'amélioration du temps de calcul, l'écart entre les temps de résolution sans et avec coupes n'est que de quelques secondes. Le fait de ne pas réduire le temps de résolution pour toutes les instances testées est en grande partie dû à l'ajout de la variable UR_{max} (taux maximal d'utilisation des véhicules) dans la fonction économique. Celle-ci est une variable continue prenant des valeurs dans l'intervalle [0,1], ce qui rend difficile une amélioration rapide de la solution.
Suite à ces constats, toujours afin d'obtenir des résultats dans un temps de calcul raisonnable, nous avons arrêté la résolution du modèle lorsqu'une solution entière se trouvant au plus à 0,5% de la solution optimale est obtenue. Nous pouvons remarquer que la valeur du temps maximal de couverture n'est pas affectée, l'impact serait donc visible sur le taux d'utilisation des véhicules. Du point de vue de la résolution tactique, cette condition pourrait affecter la comparaison des résultats avec le modèle **VP1**. Du point de vue de la solution (i.e. nombre de véhicules affectés aux zones), elle ne devrait pas vraiment avoir d'impact lors de l'analyse opérationnelle (e.g. couplage avec le modèle de simulation). Comme nous le verrons au chapitre suivant, ceci est vrai, en particulier, si une bonne politique de gestion opérationnelle des véhicules est implémentée.

Pour analyser ces indicateurs, les résultats des tests numériques sont présentés au tableau 6.7. Avec cette formulation, l'intérêt du modèle est de trouver une configuration de zones de telle façon que le temps de couverture maximal (T) et le taux maximum d'utilisation des véhicules (UR_{max}) soient tous les deux minimisés. Comme expliqué précédemment, le but d'ajouter UR_{max} dans la fonction économique est de mieux équilibrer la charge des véhicules lors de leur affectation aux zones créées. En effet, comme illustré par les résultats, nous arrivons à obtenir un taux d'utilisation moyen qui est toujours acceptable (environ 71%), alors que son écart-type est nettement diminué. Cette diminution est en moyenne de 73%, selon le nombre de zones et la taille de la flotte de véhicules. Ceci est aussi visible à partir de l'écart entre les valeurs maximale et minimale du taux d'utilisation des véhicules. Remarquons aussi que la distribution de la demande entre les zones reste équivalente, par rapport à sa valeur moyenne et à son écart-type.

Tab. 6.6 – Impact des coupes sur les temps de calcul (en sec.) pour le modèle **VP2**

Nombre de véhicules	24			30		
Nombre de zones	5	10	15	5	10	15
Sans coupes	421*	822	100*	225*	1181*	215
Avec (6.12)	665	1287	624	1434	2410	266
Avec (6.13)	665	744	116	548	1717	268
Avec (6.14)	479	705*	195	655	1218	148*
Avec (6.15)	805	8150	48515	582	9607	18321

*Meilleurs temps

Tab. 6.7 – Résultats du modèle **VP2**

Nombre de véhicules : 24			
Nombre de zones maximum	5	10	15
Nombre de zones créées	5	10	15
Temps maximal de couverture	73	50	41
Utilisation moyenne des véhicules	72,6%	72,1%	71,9%
Ecart-type de l'utilisation des véhicules	00,2%	04,77%	12,2%
Utilisation minimale des véhicules	72,3%	61,6%	52,4%
Utilisation maximale des véhicules	72,9%	76,7%	89,1%
Demande moyenne des zones	12526,8	6263,4	4175,6
Ecart-type de la demande des zones	8110,3	4109	2044,7
Nombre de véhicules : 30			
Nombre de zones maximum	5	10	15
Nombre de zones créées	5	10	15
Temps maximal de couverture	73	50	41
Utilisation moyenne des véhicules	59,2%	57,5%	57,7%
Ecart-type de l'utilisation des véhicules	3,0%	5,4%	5,6%
Utilisation minimale des véhicules	56,7%	44,8%	40,1%
Utilisation maximale des véhicules	63,7%	61,7%	61,8%
Demande moyenne des zones	12526,8	6263,4	4175,6
Ecart-type de la demande des zones	8382,7	3346,9	1898,6

La stratégie de création de zones et d'affectation des véhicules en prenant en compte à la fois le temps de couverture et le taux maximal d'utilisation de véhicules donne bien des avantages par rapport à la distribution de la charge de travail (demandes de transport) des véhicules dans les zones. Nous pouvons donc croire à une meilleure satisfaction dynamique des requêtes au niveau opérationnel de gestion.

6.2.2.3 Prise en compte de la demande des zones

Comme expliqué précédemment, le modèle **VP2** permet de prendre en compte à la fois le temps de couverture et le taux d'utilisation des véhicules. Le nombre de véhicules affecté à chaque zone doit être suffisant pour satisfaire la demande, mais entre les zones elles-mêmes, il peut y avoir des différences importantes par rapport à la demande réelle [89]. Or, nous pourrions être intéressés par une distribution plus uniforme de la demande entre les zones. Pour ce faire, nous pouvons définir une variable D_{max} comme étant la demande maximale supportée par une zone, qui est ensuite ajoutée à la fonction objectif avec un poids CO_3. Nous devons aussi ajouter les containtes (6.30) à notre formulation afin de tenir compte de cette demande maximale par rapport aux demandes associées à chacune des zones.

$$D_{max} \geq \sum_{k \in K} D_k Y_{jk} \qquad \forall j \in J \tag{6.30}$$

Une autre façon d'équilibrer la demande entre les zones est de le faire indirectement en utilisant l'équation (6.11). Celle-ci nous permet d'avoir une affectation plus uniforme des chariots aux zones et, puisque leur capacité (en termes de disponibilité par unité de temps) est connue, nous obtenons également un meilleur équilibrage de la demande. Notre intérêt

est donc de considérer un nombre maximum de véhicules dans les zones. Pour ce faire, nous avons besoin de définir une variable V_{max} représentant ce nombre maximum de véhicules. Cette variable remplacera D_{max} dans la fonction économique. Du point de vue conceptuel, la prise en compte de la variable D_{max} dans la fonction objectif permet d'évaluer directement la demande maximale des zones. Cependant, du point de vue de la résolution du programme, il est préférable d'utiliser V_{max} car, puisqu'elle prend des valeurs discrètes, l'impact d'un changement de sa valeur est plus marqué qu'avec les valeurs continues de la variable D_{max}.
Dans notre formulation présentée ci-dessous, nous avons donc gardé la variable V_{max} et les contraintes (6.30) ont été remplacées par les contraintes (6.40) qui tiennent compte du nombre maximum de véhicules dans une zone (et donc indirectement de la demande associée aux zones). Les autres contraintes du modèle sont identiques à celles définies précédemment.

Modèle VP3 :

$$\text{Minimiser } (CO_1.T + CO_2.UR_{max} + CO_3.V_{max}) \tag{6.31}$$

Sous :

$$T \geq \sum_{j \in J} t_{jk} Y_{jk} \qquad \forall k \in K \tag{6.32}$$

$$\sum_{j \in J} X_j \leq P \tag{6.33}$$

$$\sum_{j \in J} Y_{jk} = 1 \qquad \forall k \in K \tag{6.34}$$

$$Y_{jk} \leq X_j \qquad \forall j \in J, k \in K \tag{6.35}$$

$$\sum_{j \in J} Z_{vj} \leq 1 \qquad \forall v \in V \tag{6.36}$$

$$\sum_{k \in K} D_k Y_{jk} \leq \ VC \sum_{v \in V} \ UR_{vj} \qquad \forall j \in J \tag{6.37}$$

$$UR_{vj} \leq Z_{vj} \qquad \forall v \in V, j \in J \tag{6.38}$$

$$UR_{max} \geq \sum_{j \in J} UR_{vj} \qquad \forall v \in V \tag{6.39}$$

$$V_{max} \geq \sum_{v \in V} Z_{vj} \qquad \forall j \in J \tag{6.40}$$

$$X_j \in \{0, 1\} \qquad \forall j \in J \tag{6.41}$$

$$Y_{jk} \in \{0, 1\} \qquad \forall j \in J, k \in K \tag{6.42}$$

$$Z_{vj} \in \{0, 1\} \qquad \forall v \in V, j \in J \tag{6.43}$$

$$UR_{vj} \in [0, 1] \qquad \forall v \in V, j \in J \tag{6.44}$$

Le modèle **VP3** présenté est aussi un modèle d'optimisation multi-objectif. On pourrait donc être intéressé par la recherche des ensembles de solutions Pareto-optimales. Comme

expliqué précédemment, une façon de les obtenir est de faire varier les valeurs des coefficients CO_1, CO_2 et CO_3 de la fonction objectif. Néanmoins, notre intérêt est de formaliser le problème et de voir l'impact d'une certaine solution tactique de définition de zones sur la performance opérationnelle du système de transport et non pas de trouver les ensembles de solutions Pareto-optimales. Pour la résolution du modèle **VP3**, les valeurs des coefficients dans la fonction objectif ont donc été fixées à $CO_1 = 100$, $C0_2 = 100$ et $CO_3 = 1$, car notre choix était de donner une importance plus grande aux critères du temps de couverture (T) et taux d'utilisation des véhicules (UR_{max}), le nombre maximal de véhicules (V_{max}) dans les zones ayant un poids moins important. Nous avons déjà expliqué le choix d'avoir $CO_1 = CO_2$, lors de la présentation du modèle **VP2**. Pour sa part, le choix de CO_3 est justifié par le fait que la valeur de V_{max} est bornée par le rapport entre la capacité des véhicules (VC) et la demande maximale affectée à une zone ($V_{max} \geq \max_j \frac{\sum_j D_j}{VC}$). La valeur du paramètre VC peut être modifiée afin de tenir compte des pannes des véhicules. Dans notre étude, nous avons choisi une disponibilité de 100%, les ressources (machines et véhicules) étant considérées fiables (voir les hypothèses présentées dans les sections 5.2.1 et 6.2.1).

Résultats expérimentaux

Suite aux résultats obtenus pour l'analyse de l'impact des coupes dans la formulation des modèles précédents, nous avons effectué quelques tests pour analyser leur impact sur le temps de résolution du modèle **VP3**. Les résultats obtenus sont présentés au tableau 6.8. Comme pour le modèle **VP2**, les coupes n'améliorent pas toujours le temps de calcul. Les seuls cas où nous avons obtenu une diminution du temps de calcul sont les cas avec 10 et 15 zones : nous avons obtenu une réduction de 10% et 18%, respectivement. L'intérêt des coupes cependant concerne l'obtention de solutions réalistes quant à l'organisation des véhicules affectés aux zones, comme expliqué précédemment. Avec cette formulation, l'impact des coupes est en grande partie dû au fait que la fonction objectif doit prendre en compte le taux maximal d'utilisation des véhicules (UR_{max}). Rappelons que celle-ci est une variable continue prenant des valeurs dans l'intervalle [0,1], ce qui rend difficile une amélioration rapide de la solution.

Tab. 6.8 – Impact des coupes sur les temps de calcul (en sec.) pour le modèle **VP3**

Nombre de véhicules		30	
Nombre de zones	5	10	15
Sans coupes	1187*	7496	455
Avec (6.12)	5235	26739	1674
Avec (6.13)	2024	9193	762
Avec (6.14)	1262	6730*	371*
Avec (6.15)	3774	22097	13138

*Meilleurs temps

Après cette analyse préliminaire, nous avons réalisé une étude expérimentale du modèle **VP3** en prenant les mêmes conditions de test que pour le modèle **VP2**. Comme expliqué précédemment, **VP3** est aussi un modèle multi-objectif. Afin de trouver les ensembles de solution Pareto-optimales, nous pouvons faire varier les valeurs des coefficients CO_1, CO_2 et $C0_3$ de la fonction objectif. Cependant, à partir des explications déjà donneés,

nous avons fixé les valeurs de ces coefficients comme $CO_1 = CO_2 = 100$ et $CO_3 = 1$. La résolution de ce modèle a aussi été arrêtée lorsqu'une solution entière se trouvant au plus à 0,5% de la solution optimale est obtenue. Les résultats des tests numériques sont présentés au tableau 6.9.

TAB. 6.9 – Résultats du modèle **VP3**

Nombre de véhicules : 24			
Nombre de zones maximum	5	10	15
Nombre de zones créées	5	10	15
Temps maximal de couverture	73	50	41
Utilisation moyenne des véhicules	73,8%	70,9%	71,5%
Ecart-type de l'utilisation des véhicules	17,6%	11,9%	9,2%
Utilisation minimale des véhicules	43,7%	46,3%	49,9%
Utilisation maximale des véhicules	89,8%	80,7%	76,9%
Demande moyenne des zones	12905,4	6263,4	4175,6
Ecart-type de la demande des zones	5014,7	3677,5	2335,3
Nombre de véhicules : 30			
Nombre de zones maximum	5	10	15
Nombre de zones créées	5	10	15
Temps maximal de couverture	73	50	41
Utilisation moyenne des véhicules	57,7%	57,4%	57,7%
Ecart-type de l'utilisation des véhicules	6,3%	2,2%	7,5%
Utilisation minimale des véhicules	52,1%	55%	36,8%
Utilisation maximale des véhicules	68,1%	60,3%	64,5%
Demande moyenne des zones	12526,8	6267	4175,6
Ecart-type de la demande des zones	5047,7	3117,6	1927,1

Comme expliqué précédemment, le but du modèle **VP3** est de trouver une configuration de zones de telle façon que le temps de couverture maximal (T), le taux maximum d'utilisation des véhicules (UR_{max}) et la demande maximale des zones (exprimée en termes du nombre maximal de véhicules dans les zones V_{max}) soient tous les trois minimisés. Cette fonction économique permet un équilibrage à la fois de la charge des véhicules et de la demande lors de leur affectation aux zones créées. En effet, comme illustré par les résultats, nous arrivons à obtenir un taux d'utilisation moyen qui est toujours acceptable : environ 70% pour une flotte de 24 véhicules, et environ 57% avec 30 véhicules. Cette diminution du taux moyen d'utilisation avec 30 véhicules s'explique par le fait que bien moins de 30 véhicules sont nécessaires pour satisfaire dynamiquement les demandes dans le réseau, comme illustré lors de notre étude par simulation. Cette diminution n'était pas visible avec les modèles **VP1** et **VP2** parce que la création de zones ne prenait pas en compte la distribution de la demande. Maintenant, avec la formulation **VP3**, si bien que la valeur de la demande moyenne demeure équivalente, la prise en compte de celle-ci dans la fonction économique permet de diminuer considérablement son écart-type, spécialement lorsque peu de zones vont être créées (une petite valeur de P) : environ 39% pour $P = 5$ zones ou encore près de 10% pour $P = 10$, la valeur avec $P = 15$ zones restant équivalent par rapport aux modèles précédents.

6.3 Extensions des modèles

Les modèles mathématiques présentés ci-avant ont été inspirés d'un cas industriel provenant de l'industrie des semi-conducteurs. Dans cette section, nous présentons une discussion des hypothèses des modèles, ainsi qu'une façon d'aborder le problème de définition des zones différemment. Nous allons nous limiter à une présentation de la façon dont les modèles peuvent être étendus.

Etant donné les pratiques courantes de gestion, les hypothèses présentées dans la section 6.2.1, telles que nous les avons décrites, sont tout à fait réalistes quant à leur application dans le cadre de la fabrication de semi-conducteurs. Aussi, elles sont applicables dans un cadre général d'étude d'un système de fabrication automatisé avec transport interne. Puisque les modèles mathématiques partent d'un agencement de machines et d'une configuration du réseau donnés, remarquons que seule l'hypothèse concernant la vitesse identique des véhicules constitue réellement une contrainte quant à la généralisation des modèles. En effet, si nous ne considérons pas de véhicules avec les mêmes vitesse, les temps de transport entre les points du réseau deviennent donc dépendants du type de véhicule transportant l'entité. Nos modèles peuvent être modifiés en considérant, soit un temps de transport entre les points i et k dépendant du véhicule v (i.e. t_{ikv}), soit en prenant un temps moyen de transport (t_{ik}) entre les points i et k du réseau et une disponibilité VC_v dépendant du véhicule v.

Une autre extension des modèles est liée à la façon dont les zones ont été créées. Pour la définition des zones, nous avons utilisé la notion de "centre de zone", laquelle a été inspirée des concepts classiques du problème de localisation d'installations. Une fois un centre de zone sélectionné (resp. une installation ouverte dans les problèmes classiques de localisation), les points appartenant à la zone sont ensuite choisis (resp. les clients à livrer) de façon à ce que le temps maximal entre le centre et le point le plus éloigné soit minimisé (resp. optimisation de la fonction économique). Si nous considérons comme objectif la minimisation de la distance maximale entre les deux points les plus éloignés appartenant à une zone, nous pourrions créer des zones plus compactes. Ainsi, la notion de "centre de zone" n'est plus utilisée et nos modèles devront être modifiés en conséquence. Pour illustrer ces modifications, reprenons notre modèle **VP1**. La fonction objectif deviendrait donc la minimisation de cette distance, que nous allons dénoter D. Ainsi, la variable X_j et la contrainte (6.3) disparaîtraient, et les contraintes (6.2) et (6.4) seraient, respectivement, remplacées par les ensembles de contraintes (6.45) et (6.46) présentées ci-après.

$$D \geq t_{ik} \times (Y_{ij} + Y_{kj} - 1) \qquad \forall i \in K, \forall k \in K \tag{6.45}$$

$$\sum_{j=1}^{P} Y_{jk} = 1 \qquad \forall k \in K \tag{6.46}$$

Remarquons enfin que la formulation de contraintes de type (6.45) pourrait être un inconvénient au moment de la résolution du programme avec un logiciel standard, à cause du grand nombre de contraintes (6.45) et du terme $(Y_{ij} + Y_{kj} - 1)$ qui donne des relaxations linéaires de mauvaise qualité.

6.4 Une méthode heuristique

Malgré l'ajout des coupes (6.12), (6.13) et (6.15), nos études expérimentales ont montré que la résolution optimale du problème est très gourmande en temps de calcul, pour un nombre raisonnable de centres de zones potentiels et de véhicules (i.e. $|J| = 30$ et $|V| \leq 30$) [89]. Comme indiqué précédemment, ces valeurs pour les ensembles J et V ont été déterminées en raison de contraintes informatiques (capacité de mémoire). L'idée étant de pouvoir prendre en compte tous les points du réseau, notre intérêt est alors de trouver une méthode heuristique capable de trouver une bonne solution approchée en un temps de calcul raisonnable. De plus, la résolution du problème tactique de positionnement fait partie d'une démarche itérative d'aide à la décision, le modèle devant être résolu plusieurs fois en le couplant avec le modèle de simulation de la fab (voir chapitre 7). Cette section présente donc un algorithme heuristique GLOUTON pour résoudre de façon efficace le problème tactique de définition de zones et de positionnement des véhicules.

6.4.1 Algorithme GLOUTON

Le but est de trouver une configuration de zones satisfaisant les contraintes du problème et en atteignant les objectifs visés. Lors de la résolution d'un problème d'optimisation, la construction d'une solution se fait souvent de manière séquentielle, l'algorithme faisant à chaque étape un certain nombre de choix. Le principe glouton consiste à faire le choix qui semble le meilleur sur le moment (choix local), sans se préoccuper des conséquences dans l'avenir, et sans revenir en arrière. Un algorithme glouton est donc un algorithme qui se dirige le plus rapidement possible vers une solution, sans la remettre en question. En raison de son "appétit démesuré", l'algorithme GLOUTON n'est pas sûr de toujours arriver à une solution optimale, mais il fournit un résultat rapidement. Même si la solution n'est pas optimale, il n'est pas rare que l'on s'en contente, notamment pour la résolution des problèmes $\mathcal{NP}$-complets [56].

La méthode gloutonne proposée est présentée dans la figure 6.2. Lorsqu'aucun des trois critères n'est amélioré dans l'une des étapes de l'algorithme, on élimine à tour de rôle les centres existants qui sont remplacés par les autres centres potentiels de zone pour voir s'il y a une amélioration possible des critères dans l'ordre de leurs priorités. Si tel est le cas, on garde cette nouvelle configuration et on continue notre programme en cherchant à augmenter le nombre de zones. Sinon, on arrête le programme. D'autre part, lorsqu'on introduit un nouveau centre de zone on réarrange la configuration des zones de façon á réduire la demande maximale ou à rendre compatible le nombre de véhicules afin de satisfaire la demandes dans les zones.

6.4.2 Résultats numériques

Dans cette section, nous allons présenter l'étude comparative des méthodes proposées. Pour cette étude, tous les points du réseau de transport sont considérés comme centres potentiels de zones (i.e. $|J| = |K| > 200$ points). Le nombre de véhicules considéré est 24 et 30, et le nombre de zones possibles est compris entre 5 et 15. Les résultats expérimentaux sont présentés dans le tableau 6.10.

Algorithme GLOUTON	
Entrées :	Nombre P de zones à trouver Ensemble K des points du réseau Temps t_{ik} entre les points i et k du réseau Nombre $\|V\|$ de véhicules Capacité VC des véhicules
Sorties :	Ensemble de zones Affectation des véhicules aux zones
Objectifs :	Critère 1 : Temps maximal de couverture Critère 2 : Demande maximale des zones Critère 3 : Dispersion du taux d'utilisation des véhicules
Etape 1 :	Choisir le centre qui garantit le minimum temps de couverture maximale.
Etape 2 :	Choisir un deuxième centre (le meilleur) qui forme avec le premier, 2 zones minimisant le temps de couverture maximale au plus égale à celui de l'étape 1, et qui améliore le critère 1, sinon qui améliore le critère 2 pour la même valeur du critère 1, qui améliore le critère 3 pour les mêmes valeurs des critères 1 et 2.
Etape 3 :	Choisir un troisième centre (le meilleur) qui forme avec les deux autres, trois zones donnant un meilleur résultat au sens du critère 1, sinon au sens critère 2, sinon au sens du critère 3.
Etape 4 :	Répéter jusqu'à ce qu'il n'y ait plus d'amélioration possible ou que l'une des contraintes ne puisse plus être satisfaite (e.g. le nombre maximum de zones).

FIG. 6.2 – Description de l'algorithme GLOUTON

Dans le tableau, nous pouvons observer une diminution de la valeur solution de T (temps de couverture maximal) lorsque le nombre P de zones à ouvrir augmente. Ceci est bien observé aussi lors de la résolution des programmes linéaires. Ce résultat est logique quant à l'objectif d'ouvrir le plus de zones possible afin de diminuer la valeur de T. La figure 6.3 présente une comparaison entre les valeurs solution de T données par les modèles mathématiques (PLNE) et l'algorithme GLOUTON. Nous pouvons remarquer que les valeurs données par la méthode gloutonne correspondent aussi aux valeurs optimales des PLNE pour les configuration avec 10 zones et 24 et 30 véhicules et pour 15 zones et 24 véhicules. Le fait qu'avec 5 zones la méthode gloutonne donne une solution de T meilleure que les modèles mathématiques s'explique par la sensibilité des PLNE par rapport aux points appartenant à l'ensemble J (points choisis qui peuvent être centre de zone), comme indiqué précédemment (voir tableau 6.2) [89]. L'impact de l'ensemble J sur la solution des PLNE a été expliqué dans la section 6.2.2.

Comme expliqué précédemment, l'algorithme GLOUTON considère que tous les points du réseau comme centres potentiels de zone (i.e. $J = K$). Nous avons donc élargi les tests et comparé les valeurs de T lorsqu'un nombre plus grand de zones est ouvert (plus de 15

TAB. 6.10 – Résultats de l'algorithme glouton

Nombre de véhicules : 24			
Nombre de zones maximum	5	10	15
Nombre de zones créées	5	10	15
Temps maximal de couverture	71	51	41
Utilisation moyenne de véhicules	72,4%	72,4%	72,4%
Ecart-type de l'utilisation des véhicules	6,52%	10,9%	15,4%
Utilisation minimale de véhicules	62,7%	50,1%	46,7%
Utilisation maximale de véhicules	79,2%	87,1%	91,7%
Demanade moyenne des zones	12518,2	6259,1	4172,7
Ecart-type de la demande des zones	11126,9	2258,9	1693,5
Temps de calcul (sec.)	115	145	188
Nombre de véhicules : 30			
Nombre de zones maximum	5	10	15
Nombre de zones créées	5	10	15
Temps maximal de couverture	71	51	42
Utilisation moyenne de véhicules	57,9%	57,9%	57,9%
Ecart-type de l'utilisation des véhicules	5,03%	3,36%	3,64%
Utilisation minimale de véhicules	50,1%	52%	47,3%
Utilisation maximale de véhicules	62,2%	62,2%	64,2%
Demanade moyenne des zones	12518,2	6259,1	4172,7
Ecart-type de la demande des zones	10847,5	2618,4	1394,2
Temps de calcul (sec.)	234	671	1420

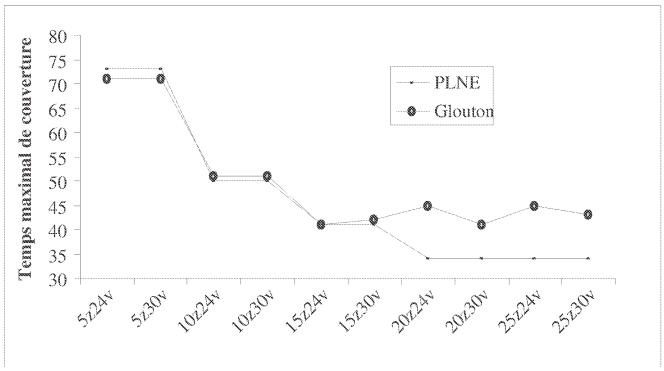


FIG. 6.3 – Temps maximal de couverture : PLNE *versus* Glouton

zones). Nous pouvons remarquer que la solution gloutonne pour le temps de couverture est plus élevée que la solution donnée par la programmation mathématique. Ce résultat est à espérer à cause de l'objectif de la méthode gloutonne qui est de trouver *une* solution faisable le plus rapidement possible. Par conséquent, cette vision "myope" ne lui permet pas de se remettre en question et, ainsi aveuglée, les choix sont faits en améliorant de manière locale, ce qui bien évidemment n'est pas forcement une bonne alternative du point de vue global du réseau.

Les autres résultats de l'algorithme GLOUTON sont présentés au tableau 6.10. Malgré ces différences de la valeur solution de T, l'algorithme GLOUTON donne des résultats accep-

tables pour la valeur moyenne du taux d'utilisation des véhicules, qui sont comparable aux résultats obtenus avec les programmes linéaires. Cela nous permet de gagner en temps de calcul et ainsi envisager la possibilité d'implémenter cet algorithme dans un système d'aide à la décision permettant la définition ou re-définition des zones et l'affectation des véhicules en temps réel pendant la période de production. Cet aspect sera discuté dans les perspectives de cette thèse. Nous pourrions imaginer aussi une méthode hybride GLOUTON-PLNE qui utiliserait la méthode gloutonne pour définir l'ensemble J de points potentiellement centre de zone (avec $|J| = 30$ points) et qui ensuite trouverait la solution optimale pour la configuration souhaitée. La méthode gloutonne peut aussi servir de base pour le développement de méthodes méta-heuristiques plus rapides et efficaces. Nous reviendrons également sur ces points dans les perspectives de cette thèse.

6.5 Conclusion

Dans ce chapitre, nous avons présenté une méthodologie formelle pour le découpage en zones du réseau de transport. Le problème a été traité du point de vue tactique. Nous avons modélisé le problème à l'aide de programmes linéaires en nombre entiers en formulant des hypothèses réalistes et en utilisant des données agrégées fournies par l'entreprise partenaire de ce projet de recherche. Nous avons également discuté la pertinence des trois modèles mathématiques proposés. Dans un premier modèle, nous avons formulé le problème afin de minimiser le temps de couverture maximal d'un point à partir du centre de la zone. Ce modèle présente des avantages quant au découpage du réseau, mais l'affectation des véhicules est réalisée afin de satisfaire la demande sans tenir compte de la charge de travail affectée à chaque véhicule (sous certains paramètres, il y aura des véhicules utilisés à presque 100% alors que d'autres le seront à moins de 20%). Nous avons donc proposé un deuxième modèle dans lequel le taux d'utilisation des véhicules est pris en compte dans la fonction objectif. Le but est ainsi d'équilibrer la charge de travail de la flotte de chariots. Le dernier modèle mathématique proposé cherche à la fois à minimiser le temps de couverture et le taux maximal d'utilisation des véhicules et, en plus, à minimiser la demande maximale des zones. Afin de faciliter la résolution du programme dans un logiciel standard de programmation linéaire, ce dernier objectif est atteint de façon indirecte en minimisant le nombre maximal de véhicules dans les zones. La capacité des véhicules étant connue et identique pour tous les chariots, le lien entre la demande et le nombre de véhicules permet effectivement d'équilibrer la demande entre les zones créées. A la fin du chapitre, nous avons d'une part discuté des extensions possibles des modèles proposés et d'autre part présenté une méthode heuristique permettant de réduire le temps de résolution, et qui, potentiellement, peut prendre en compte des contraintes et des relations non-linéaires entre les variables. Un atout très important de la méthode gloutonne se trouve dans le fait qu'elle est capable de choisir les centres des zones à partir de tous les points du réseau, contrairement aux modèles mathématiques qui partent d'un sous-ensemble J de points qui peuvent devenir centre de zones. Dans notre étude expérimentale, $|J| = 30$. Un inconvénient de cet ensemble J concerne la sensibilité de la solution fournie par les PLNE, notamment en termes de la valeur de la fonction objectif.

Comme expliqué précédemment, la solution fournie par ces modèles d'optimisation constitue le niveau tactique, c'est-à-dire, la planification des ressources de transport. La mise en place d'une démarche itérative d'aide à la décision nécessite le couplage de la solution

au niveau tactique avec le modèle de simulation de la fab. Dans le chapitre suivant, nous allons présenter et comparer plusieurs politiques de gestion opérationnelle du système de transport, afin d'analyser son comportement dynamique et l'impact des solutions tactiques sur les performances de la fab.

Chapitre 7

Gestion opérationnelle des véhicules

Nous avons vu lors de l'analyse du cas industriel au chapitre 5 qu'un découpage du réseau de transport en zones de gestion permettrait d'améliorer les performances de l'AMHS et de la fab. Au chapitre 6, nous avons présenté une méthodologie formelle pour réaliser ce découpage à partir d'une analyse tactique. Dans ce chapitre, nous allons récupérer les résultats de cette optimisation tactique pour les incorporer dans le modèle de simulation de la fab. Afin de prendre en compte l'évolution dynamique du système, plusieurs politiques opérationnelles de gestion des véhicules peuvent être mises en place. Nous allons analyser certaines d'entre elles. Nous commençons par un rappel du problème, suivi de la description des politiques que nous allons considérer et leur analyse avec la simulation. A la fin du chapitre, nous réalisons une analyse plus étendue avec la politique de gestion donnant les meilleures performances, et la comparons avec les résultats d'une gestion du transport sans zones. Une étude sur la cohérence entre les niveaux tactique et opérationnel est aussi présentée.

7.1 Rappel de la problématique opérationnelle

Comme nous l'avons expliqué au chapitre 3, la gestion opérationnelle d'un système automatisé de transport englobe les approches et algorithmes pour contrôler et coordonner de manière automatique les ressources du niveau matériel (e.g. véhicules) de l'AMHS global, de sorte que ce dernier fonctionne de manière harmonieuse et efficace. Comme nous l'avons également expliqué précédemment (voir section 3.3), les problèmes qui concernent la partie opérationnelle du système de transport sont le dispatching, le routage et le positionnement des véhicules, ainsi que la recherche de tâches de transport dans le réseau. Rappelons brièvement que le problème de dispatching des véhicules est en fait un problème d'ordonnancement où l'objectif est de mettre en place une règle ou des règles permettant d'affecter les véhicules libres aux demandes de transport en attente. Une fois qu'un véhicule est affecté à une mission de transport, le problème de routage consiste alors à trouver un trajet réalisable qui le mène de sa position actuelle à sa destination finale. Le problème de recherche de tâches consiste à définir l'espace de couverture dans lequel les véhicules, une fois disponibles, vont chercher les entités demandant un service de transport. Ce problème est lié à la fois au problème de dispatching et au problème de positionnement des véhicules. Ce dernier, comme nous l'avons vu au chapitre précédent, consiste à découper le réseau de transport en zones de gestion et d'affecter une flotte de véhicules de telle façon que le système soit couvert (i.e. qu'il y ait des véhicules proches des points de demande) et que

les demandes de transport soient satisfaites dans les meilleurs délais. Comme nous l'avons vu, cette approche du problème de positionnement constitue le niveau tactique. Au niveau opérationnel, nous devons implémenter une ou des politiques pour la gestion des véhicules dans le réseau découpé en zones.

Rappelons que nous partons d'une configuration du réseau (i.e. localisation des points de chargement/déchargement) qui est donnée. Le découpage du réseau en zones et la définition des points appartenant à chaque zone sont déterminés par la solution du problème tactique. Rappelons également que cette solution tactique considère que le service des requêtes est réalisé avec des véhicules stationnés au centre de la zone. Or, au niveau opérationnel, nous devons considérer les déplacements des véhicules dans leur zone et entre les zones. Lorsque les véhicules restent à l'intérieur de la zone à laquelle ils ont été affectés, le service des demandes de transport est effectué en suivant la stratégie de recherche de tâches et les règles de dispatching mises en place. Si les véhicules changent de zone, nous devons définir des politiques de gestion afin de décider comment ceux-ci vont satisfaire les demandes de transport dans chacune des zones. En particulier, nous allons devoir répondre à des questions du type :

- Que faire avec un véhicule une fois qu'il est sorti de sa zone ? Doit-il revenir à sa zone d'origine ? Est-il autorisé à rester dans la zone d'arrivée ?
- Lorsqu'il n'y a plus de demandes dans une zone, un véhicule libre a-t-il le droit de servir les demandes d'une autre zone ? Dans ce cas, comment fait-on le choix de la zone à servir ?
- Est-il possible d'avoir un certain nombre de véhicules non affectés aux zones ? Quel est leur impact sur les performances du système ?

Dans les sections qui suivent, nous allons proposer des stratégies pour répondre aux questions précédemment posées. La section 7.2 présente plusieurs politiques pour la gestion opérationnelle des véhicules dans un réseau de transport découpé en zones. Dans la section 7.3, nous étudions de façon approfondie les performances de la politique la plus performante parmi celles présentées. Cette analyse est effectuée en deux parties. Premièrement, la solution de la planification tactique est implémentée dans le modèle de simulation de la fab. Deuxièmement, nous allons tester la cohérence entre les niveaux de décision tactique (à travers l'optimisation) et opérationnel (à l'aide de la simulation). Enfin, la section 7.4 présente une étude sur l'intérêt d'avoir une partie de la flotte de véhicules non affectés aux zones créées au niveau de la planification tactique. Les performances en simulation sont testées et comparées avec les autres scénarii étudiés.

7.2 Politiques de gestion des zones et des véhicules

La distribution des véhicules donnée par le niveau tactique est réalisée à partir d'une estimation de la demande de transport de chaque point du réseau. Au cours de la période de production, les demandes dans les zones peuvent varier. Or, les besoins effectifs en nombre de véhicules varient aussi en fonction de la demande. Puisque les estimations du niveau tactique sont des valeurs moyennes, des politiques de gestion opérationnelle autorisant la redistribution des véhicules entre les zones ou le choix d'avoir une flotte de véhicules non pré-affectés aux zones pourraient être mises en place afin de faire face à ces variations de la demande.

Nous présentons ci-après plusieurs politiques de gestion opérationnelle. Une étude par

simulation est réalisée dans le but d'analyser leurs performances et leur impact sur les indicateurs de la fab. Toutes ces politiques ont en commun le découpage en zones du réseau de transport, ainsi que l'affectation des véhicules. La différence fondamentale entre elles concerne le critère utilisé pour l'évolution dynamique de l'affectation de véhicules (e.g. selon le besoin en véhicules, selon l'évolution de la demande de zones, etc.). Les zones, elles aussi, peuvent être gérées et redéfinies tout au long de l'horizon de production. Dans cette thèse, une fois les zones créées et une affectation initiale de chariots réalisée, nous nous concentrons uniquement sur la façon dont les véhicules sont gérés dans le système.

Le tableau 7.1 résume les différentes caractéristiques des politiques proposées. Nous pouvons nous apercevoir de leur caractéristiques communes précédemment enoncées. Les autres conditions d'opération des politiques varient selon la stratégie à mettre en place. En particulier, elles se différencient par les critères de ré-affectation dynamique des véhicules aux zones et par l'autorisation donnée aux véhicules de changer ou non de zone selon les besoins globaux ou locaux dans le système. La description détaillée de ces politiques est présentée dans les sections qui suivent. Les tests de simulation comparatifs ont été réalisés sur le modèle de simulation de la fab que nous avons présentée au chapitre 5. Rappelons que la gamme opératoire comporte environ 700 étapes avec des passages multiples sur la même séquence de postes (en particulier, 40 fois sur l'étape de photolithographie). Nous considérons un seul type de produit et la même priorité d'exécution pour tous les lots. Les files d'attente avant et après les machines sont gérées selon l'ordre d'arrivée dans le stock (FIFO, *First In First Out*). Dans la fab, il y a plus de 200 machines. Leur agencement et leur configuration du réseau de transport sont déjà définis.

Tab. 7.1 – Comparaison des politiques opérationnelles

	Politique opérationnelle			
Caractéristiques	**ZVBRet**	**ZVBProp**	**ZVBDem**	**ZVBDemFab**
Découpage du réseau en zones	*	*	*	*
Affectation initiale d'une flotte de véhicules aux zones	*	*	*	*
Ré-affectation des véhicules :				
-Pas de ré-affectation	*			
-Selon les besoins en véhicules		*		
-Selon la demande des zones			*	
-Selon la demande dans la fab				*
A l'arrivée dans une nouvelle zone, un véhicule :				
-ne change pas de zone	*			
-peut changer de zone seulement si la flotte actuelle est inférieure à la flotte requise		*	*	*

Rappelons aussi que le modèle de simulation de la fab a été programmé sous AutoMod^{TM}. La mise en place des règles de dispatching (sélection et affectation des véhicules aux demandes de transport) et des politiques opérationnelles dans les zones est effectuée à partir des paramètres standards du logiciel, ainsi qu'à l'aide de fonctions et de procédures de

programmation en langage propre à AutoMod^{TM}. Suite au développement informatique, la principale limitation pratique concerne la manière avec laquelle les véhicules reçoivent les ordres de transport. Dans notre modèle informatique, une fois qu'un véhicule est dispatché pour exécuter une tâche (i.e. un déplacement), il se doit de la terminer avant de pouvoir effectuer une autre tâche (i.e. les tâches ne sont pas préemptives). Ceci n'est pas contraignant lorsque la tâche à réaliser par le véhicule est un déplacement en charge. Du point de vue pratique, le problème survient lorsqu'il s'agit des déplacements à vide. Un chariot exécute un déplacement à vide :

1. S'il doit aller chercher une entité pour ensuite effectuer un déplacement en charge,
2. S'il n'y a pas de demandes en attente de transport au moment de la prise de décision et que ce véhicule va vers le centre de la zone.

D'un point de vue pratique, cette implémentation pourrait avoir les limitations suivantes. Dans le premier cas, le changement de route d'un véhicule allant chercher une entité peut être bénéfique si des lots prioritaires (*hot lots, super hot lots*,...) sont aussi dans le système. Notre modèle de la fab cependant ne considère pas les lots prioritaires.
Dans la seconde situation, le chariot reste en circulation à l'intérieur de la zone à laquelle il appartient. Ce circuit est déterminé par deux points de la zone dont un est son centre. Le fait que les tâches soient non préemptives oblige le chariot à réaliser un déplacement complet entre les deux points avant de se voir affecter une autre tâche (déplacement à vide pour chercher une entité). Dans le système de pilotage réel, le véhicule pourrait être affecté à cette tâche avant de compléter le circuit.

Rappel de la notation utilisée

Dans les descriptions des politiques opérationnelles pour la gestion des véhicules, nous allons faire le lien avec les décisions prises au niveau tactique. Pour mieux comprendre ces liens, nous rappelons brièvement la notation utilisée pour la définition des variables de notre modèle de découpage et positionnement tactique des véhicules :

- X_j : variable binaire qui vaut 1 si le point j est choisi comme centre d'une zone, 0 sinon,
- Y_{jk} : variable binaire qui vaut 1 si le point k du réseau appartient à la zone dont le point j est le centre, 0 sinon,
- Z_{vj} : variable binaire qui vaut 1 si le véhicule v est affecté à la zone dont le point j est le centre, 0 sinon.

Nous avons maintenant tous les éléments nécessaires pour présenter et analyser les politiques opérationnelles mises en œuvre pour la gestion dynamique des véhicules dans un réseau de transport découpé en zones. Chaque politique est présentée de façon indépendente. Nous donnons les principaux éléments qui nous ont motivé à les proposer, ainsi que leur description formelle (en faisant le lien avec les variables de décision du niveau tactique). Des tests de simulation sont aussi développés afin de quantifier leur impact sur les performances de la fab. Ces tests ont été réalisés pour les règles de dispatching STD (*Shortest Travel Distance*) et CL-FCFS (*Current Location-First Come First Serve*) et uniquement pour une configuration avec 5 zones et 24 véhicules. Rappelons brièvement que STD choisit toujours l'entité demandeuse la plus proche du véhicule, alors que CL-FCFS regarde d'abord s'il y a une demande à la localisation actuelle du véhicule et ensuite choisit les entités en attente de transport selon l'ordre dans lequel elles ont fait la demande. La

période simulée a été de 4 mois, dont 2 mois considérés comme période transitoire (les statistiques obtenues pendant cette période transitoire n'ont pas été prises en compte pour l'analyse des résultats).

7.2.1 Politique ZVBRet : Définition de zones, nombre de véhicules borné par zone/retour du véhicule

Motivation

Le but de cette première politique est de conserver pendant toute la simulation l'affectation des véhicules aux zones donnée par la planification tactique. Cette configuration reste fixée tout au long de l'horizon de production. Cette politique utilise donc une stratégie de gestion des chariots statique et aveugle. En raison de l'occurrence dynamique des événements dans le système de fabrication, *a priori* cette politique statique et aveugle ne donne pas des bonnes performances, mais elle nous permettra de comprendre le véritable comportement dynamique du réseau découpé en zones de gestion. L'analyse de résultats devra nous donner aussi des pistes pour l'amélioration de la stratégie de contrôle des véhicules.

Description

A partir des résultats du niveau tactique, le réseau de transport est divisé en un certain nombre de zones et une flotte de véhicules est affectée à chacune des zones (i.e. pour la zone dont le centre est le point j, nous devons avoir $\sum_v Z_{vj}$ véhicules). Les véhicules affectés à une zone ne peuvent servir que les demandes qui sont émises par les points appartenant à cette zone (i.e. $Y_{jk} = 1$). Sous cette politique, à chaque fois qu'un chariot v entre dans la zone dont le point j est le centre, autre que sa zone d'affectation initiale, le véhicule v revient à sa zone d'origine pour y servir les demandes. Si dans une zone il n'y a plus de demandes, alors les véhicules restent en circulation à l'intérieur de la zone afin d'être prêts pour satisfaire une requête.

Simulations

Afin d'observer l'impact de cette politique de gestion opérationnelle sur les indicateurs clés de la fab, nous l'avons implémentée dans le modèle de simulation de fab. Le tableau 7.2 présente les résultats obtenus. Les indicateurs de performance ont été définis au chapitre 5. Rappelons que les valeurs absolues du temps de cycle ne peuvent être présentées pour des raisons de confidentialité. A la place, nous présentons le X-factor qui est défini par la relation entre le temps de cycle et le temps total de process ([(*temps de cycle*)/(*temps total de process*)]). Le X-factor donne ainsi une bonne représentation du pourcentage du temps de cycle dû aux activités de non-production. L'objectif est donc d'avoir un X-factor aussi proche que possible de 1.

Nous pouvons observer que plus de 90% du temps de non-production d'une entité est dû à l'attente d'un véhicule pour effectuer une opération de transport. Cette politique n'est donc pas efficace car le taux d'utilisation des chariots est de 68% avec la règle STD ou de 82% avec la règle CL-FCFS. Rappelons qu'au chapitre 5, nous avons vu qu'un flotte de 24 véhicules n'est pas bloquante. Aussi, à cause de l'affectation statique des chariots aux zones, avec cette politique, il se peut qu'il y ait à certains endroits de la fab, des requêtes en attente d'un chariot qui devront attendre pendant une période de temps assez importante

TAB. 7.2 – Résultats des simulations : Politique ZVBRet

Nbr. véhicules = 24		
Règle de dispatching	STD	CL-FCFS
X-factor	8,25	6,92
Ecart-type du temps de cycle	10,13	7,88
Temps d'attente d'un véhicule (% du temps de non-production)	95%	97%
Utilisation maximale des machines	69%	74%
Utilisation de véhicules	68%	82%
Temps de livraison	2,56	2,33

car les véhicules disponibles ne peuvent pas les servir, ces chariots n'étant pas affectés à la zone ou aux zones demandeuses. De ce fait, le taux maximal d'utilisation des machines est aussi détérioré, environ 70%, alors que nous avons vu au chapitre 5 que ce taux pouvait facilement atteindre 93%.

7.2.2 Politique ZVBProp : Définition de zones, nombre de véhicules maximum borné par zone, ré-affectation selon le nombre actuel de véhicules

Motivation

Avec la politique ZVBRet, décrite précédemment, les véhicules étant toujours affectés à une zone selon la solution donnée au niveau tactique, le comportement dynamique de la demande n'est pas traité de façon efficace. Nous pouvons imaginer des politiques opérationnelles qui autorisent les véhicules à changer de zones, selon un critère à définir qui nous permettrait d'améliorer les indicateurs de performance de la fab et du système de transport. Or, le comportement dynamique du système de fabrication fait que les véhicules se déplacent d'une zone à une autre de la fab. Nous pouvons, à chaque prise de décision (i.e. à chaque instant qu'un véhicule se libère) considérer une stratégie de ré-affectation des véhicules libres selon le besoin en nombre de véhicules des zones. Cette comparaison est effectuée à partir des critères définis par la planification tactique. Cela nous permettra de nous assurer que chaque zone aura toujours le nombre de véhicules nécessaire.

Description

Avec cette politique, le critère d'affectation d'un véhicule libre peut changer. Le réseau de transport est toujours divisé en un certain nombre de zones, et un nombre de véhicules est affecté pour chacune des zones (i.e. pour la zone dont le centre est le point j, nous devons avoir $\sum_v Z_{vj}$ véhicules), et les véhicules affectés ne peuvent servir que les demandes provenant des points appartenant à la zone (i.e. $Y_{jk} = 1$). Cette configuration est définie au niveau tactique. Sous cette politique, nous devons nous assurer que chaque zone aura toujours le nombre approprié de véhicules disponibles pour satisfaire les requêtes (taille "tactique" de la flotte). Afin d'assurer le nombre requis de véhicules, à chaque fois qu'un chariot v entre dans la zone dont le point j est le centre, il faut demander si après livraison de l'entité, le véhicule v a le droit ou non de rester dans cette "nouvelle" zone (i.e. la zone d'arrivée). Cette condition est évaluée en comparant le nombre actuel de véhicules et la

taille nécessaire. Si le chariot a le droit d'y rester (i.e. si le nombre actuel de véhicules est strictement inférieur à $\sum_v Z_{vj}$), alors il est affecté à cette zone. Si le véhicule v ne peut pas rester, alors il est dispatché vers la zone ayant le plus petit ratio entre la taille actuelle de la flotte et la taille requise. Remarquons enfin que s'il n'y a plus de demandes dans une zone, alors les véhicules restent en circulation entre deux points à l'intérieur de la zone. Les implications de cette contrainte de circulation ont été expliquées dans la section 7.2.

Simulations

Les conditions de simulation utilisées pour le test de cette politique sont les mêmes que pour la politique précédente. Les résultats obtenus sont présentés au tableau 7.3. Nous observons maintenant une petite amélioration du temps de cycle (indiqué par le X- factor) et de sa variance avec l'utilisation de la règle de dispatching STD, ainsi que du taux d'utilisation des véhicules. Avec la règle CL-FCFS, les valeurs du temps de cycle (X-factor) et de sa variance ne sont pas améliorées par rapport à la politique ZVBRet. De plus, le pourcentage du temps de non-production dû à l'attente d'un véhicule par l'entité demeure équivalent pour les deux règles de dispatching.

Tab. 7.3 – Résultats des simulations : Politique ZVBProp

Nbr. véhicules = 24		
Règle de dispatching	STD	CL-FCFS
X-factor	8,05	6,99
Ecart-type du temps de cycle	9,76	8,73
Temps d'attente d'un véhicule (% du temps de non-production)	95%	96%
Utilisation maximale des machines	71%	73%
Utilisation de véhicules	75%	79%
Temps de livraison	2,71	2,36

Comme pour la règle précédente, ces mauvaises performances sont dues principalement au fait que les demandes en attente de transport ne sont pas traitées de façon efficace. Avec la règle ZVBRet, une fois les véhicules affectés aux zones, il n'était pas possible d'effecteur une ré-affectation pendant l'horizon de production. Maintenant, avec la règle ZVBProp, un véhicule peut rester dans une zone si celle-ci manque de véhicules. Cette ré-affectation est effectuée à partir du résultat du ratio calculé entre le nombre actuel de véhicules et le nombre requis (déterminé au niveau tactique). Cependant, encore une fois, le fait d'envoyer les véhicules libres aux zones n'ayant pas la taille requise calculée au niveau tactique ne garantit pas forcément un traitement dynamique des requêtes en attente de transport : la zone n'ayant pas la flotte requise n'est pas forcément celle qui a le plus besoin de chariots en termes de requêtes en attente. La valeur du temps cycle de l'entité peut donc être pénalisée.

7.2.3 Politique ZVBDem : Définition de zones, nombre de véhicules maximum borné par zone, ré-affectation selon la demande dans les zones

Motivation

La prise en compte des besoins en véhicules pour leur ré-affectation lorsqu'ils deviennent disponibles (politique ZVBProp) nous a permis d'améliorer un peu les indicateurs de performance. Néanmoins, ces besoins étant exprimés en fonction des véhicules par rapport à la taille définie au niveau tactique, les besoins réels donnés par les demandes de transport en attente ne sont pas considérés. La ré-affectation est toujours évaluée de manière statique. L'évolution de la demande dans le temps peut être prise en compte en comparant la quantité de lots en attente de transport dans chacune des zones. La politique ZVBDem proposée ci-après permet de ré-affecter un véhicule libre à la zone ayant le plus de demandes en attente au moment de la prise de décision.

Description

Avec cette politique, nous considérons les demandes des zones pour la ré-affectation des véhicules libres. Afin d'assurer le nombre requis de véhicules, à chaque fois qu'un chariot v entre dans la zone dont le point j est le centre, il faut demander si après livraison de l'entité, le véhicule v a le droit ou non de rester dans la zone d'arrivée. Comme dans la politique ZVBProp, cette condition est évaluée à partir du nombre actuel de véhicules et la taille nécessaire qui est donnée par le niveau tactique. Si le chariot a le droit d'y rester (i.e. si le nombre actuel de véhicules est inférieur au nombre requis), alors il est affecté à cette zone ; sinon, il est dispatché vers une autre zone. A la différence des politiques précédemment décrites, le critère de ré-affectation de ZVBDem consiste à prendre en compte les demandes de transport des zones. Ainsi, le véhicule libre est dispatché vers la zone ayant le plus de requêtes en attente de transport. Remarquons enfin que s'il n'y a plus de demandes dans la zone, alors les véhicules restent en circulation à l'intérieur de la zone.

Simulations

Les conditions de simulation utilisées pour le test de cette politique sont les mêmes que précédemment. Les résultats obtenus sont présentés au tableau 7.4. Nous observons maintenant une amélioration du temps de cycle (indiqué par le X- factor) et de sa variance, avec les deux règles de dispatching implémentées. La valeur du taux maximal d'utilisation des machines est aussi supérieure aux valeurs obtenues avec les deux politiques précédentes. En ce qui concerne les indicateurs liés à la performance du système de transport lui-même, nous observons une diminution d'environ 7% du temps moyen de livraison avec la règle STD. La valeur du taux d'utilisation des véhicules a atteint 100%.

Puisqu'avec cette politique un véhicule libre est dispatché vers la zone ayant le plus de requêtes en attente, elle donne de meilleurs résultats que les précédentes. Nous ne sommes néanmoins toujours pas complètement satisfaits car le pourcentage de temps de non-production dû à l'attente d'un chariot est de 83% avec la règle STD et de 96% avec la règle CL-FCFS. En comparant avec les résultats des simulations obtenus au chapitre 5, nous sommes encore très loin des valeurs obtenues sans le découpage du réseau en zones.

TAB. 7.4 – Résultats des simulations : Politique ZVBDem

Nbr. véhicules = 24 Règle de dispatching	STD	CL-FCFS
X-factor	5,05	4,98
Ecart-type du temps de cycle	5,25	5,33
Temps d'attente d'un véhicule (% du temps de non-production)	83%	96%
Utilisation maximale des machines	77%	76%
Utilisation de véhicules	100%	100%
Temps de livraison	2,52	2,51

Cela montre donc qu'au lieu de mettre en place des mécanismes d'amélioration locale, nous devons considérer une vision plus globale du système.

7.2.4 Politique ZVBDemFab : Définition de zones, nombre de véhicules maximum borné par zone, ré-affectation selon la demande dans la fab

Motivation

La prise en compte de la demande en attente pour la ré-affectation des véhicules libres (politique ZVBDem) nous a permis d'améliorer les performances industrielles, par rapport aux politiques qui ne considèrent pas la demande de transport (politiques ZVBRet et ZVBProp). Néanmoins, cette prise en compte de la demande est faite de manière localisée, en comparant les besoins sectorisés accumulés (quantité de lots en attente de transport). Or, le temps d'attente est un critère bien plus important lorsque l'objectif de la gestion des opérations est de réduire le temps de cycle. La politique ZVBDemFab que nous allons proposer maintenant permet de ré-affecter les véhicules libres aux zones, tout en considérant le temps d'attente. La gestion des requêtes en attente est donc effectuée de manière plus globale.

Description

Comme toutes les autres politiques déjà présentées, selon la solution du niveau tactique, la fab est divisée en un certain nombre de zones, une flotte de véhicules leur est affectée ($\sum_v Z_{vj}$) et ceux-ci ne peuvent servir que les demandes provenant des points appartenant à la zone (i.e. $Y_{jk} = 1$). Une fois un véhicule arrivé à une nouvelle zone, il faut demander s'il a le droit d'y entrer. Si la flotte actuelle dans la zone d'arrivée est suffisante, alors le véhicule est autorisé à servir une demande de transport provenant d'un point quelconque du réseau (selon la règle de dispatching mise en place). Le véhicule v sera ensuite affecté à la zone dont il va servir la requête. Avec cette stratégie, l'objectif est d'avoir une vision plus globale de la fab pour satisfaire les demandes en attente. Remarquons que les véhicules changent de zone uniquement si la zone d'arrivée a la taille requise ; autrement, les véhicules restent toujours attachés à leurs zones.

Simulations

Les conditions de simulation utilisées pour le test de cette politique sont les mêmes que précédemment. Les résultats obtenus sont présentés au tableau 7.5. Nous observons maintenant une grande amélioration du temps de cycle (indiqué par le X-factor) et de sa variance, avec les deux règles de dispatching implémentées. En particulier, nous arrivons à diminuer le temps de cycle d'un peu plus de 70% par rapport à la politique ZVBDem, et de quasiment 80% par rapport à la politique ZVBRet qui a été la première politique testée. La valeur du taux maximal d'utilisation des machines a maintenant atteint la valeur de 93%, ce qui correspond à l'utilisation maximale des ressources de production lorsque l'AMHS n'est pas la ressource goulot. En ce qui concerne les indicateurs liés à la performance du système de transport lui-même, nous observons une nette diminution du temps moyen de livraison (25% avec la règle de dispatching CL-FCFS et 28% avec la règle STD, par rapport à la politique de ZVBDem). Nous remarquons également que le système de transport n'est pas une ressource bloquante car le taux d'utilisation des véhicules est de 93% avec STD et 97% avec CL-FCFS.

Tab. 7.5 – Résultats des simulations : Politique ZVBDemFab

Nbr. véhicules = 24		
Règle de dispatching	STD	CL-FCFS
X-factor	1,33	1,42
Ecart-type du temps de cycle	0,05	0,04
Temps d'attente d'un véhicule (% du temps de non-production)	49%	60%
Utilisation maximale des machines	93%	93%
Utilisation de véhicules	93%	97%
Temps de livraison	1,80	1,89

Avec cette politique, la gestion dynamique des véhicules est réalisée en ayant une vision plus globale des demandes dans la fab. Par rapport à la politique ZVBDem qui considère également les demandes de façon dynamique, la politique ZVBDemFab fait que la réaffectation des véhicules libres est effectuée en les dispatchant vers le point du réseau ayant émis une requête (au lieu de le faire vers la zone ayant le plus de demande en attente). Cette vision globale du système permet de prendre en compte le temps effectif d'attente de l'entité, notamment avec la règle de dispatching CL-FCFS. Remarquons enfin qu'avec cette politique, la notion de zones est toujours utilisée car le dispatching d'un véhicule libre vers le point de demande est effectué seulement si la zone d'arrivée a le nombre requis de véhicules pour servir sa propre demande (dans ce cas, le véhicule arrivant est de trop et il peut donc être dispatché vers un autre point du réseau ayant émis une requête).
Ces résultats sont encourageants et nous incitent à développer une étude de simulation approfondie de l'impact de la stratégie de découpage en zones (optimisation tactique) couplée avec une gestion opérationnelle des véhicules. Celui-ci est l'objectif de la section suivante.

7.3 Couplage optimisation-simulation

Dans cette section, nous allons implémenter la solution optimale obtenue au niveau tactique sur le modèle de simulation de la fab. Nous allons utiliser la politique opérationnelle

ZVBDemFab pour analyser l'impact des modifications des paramètres tactiques sur les indicateurs de performance. Cette étude est composée de deux parties. Premièrement, nous allons mettre en place la solution de la planification tactique avec plusieurs valeurs pour la taille de la flotte de chariots et le nombre de zones. Deuxièmement, nous allons analyser la cohérence entre les deux niveaux de décision. Notre objectif est de constater qu'une mauvaise solution tactique donne aussi une mauvaise solution opérationnelle. Pour ce faire, nous allons procéder en deux étapes. D'abord, nous allons modifier la distribution des véhicules dans les zones tout en gardant la même configuration. Les solutions de simulation seront comparées avec les cas où il n'y a pas eu de modifications (solution optimale de la planification tactique). Ensuite, nous allons forcer le modèle tactique à donner de mauvaises solutions pour le temps de couverture. La configuration des zones et l'affectation des véhicules obtenues seront implémentées sur le modèle de simulation, et les performances opérationnelles seront comparées.

7.3.1 Mise en place de la solution de la planification tactique

Dans cette section, nous allons présenter les résultats obtenus lors de la mise en place de la solution de la planification tactique sur le modèle de simulation opérationnel de la fab [90]. Nous avons conçu un plan d'expérience à facteurs multiples sans répétitions [81]. Les principaux facteurs qui peuvent avoir des effets sur les indicateurs de performance sont la règle de dispatching des véhicules, la politique de gestion par zones et la taille de la flotte. Les différents niveaux pour chaque facteur dans le plan d'expérience sont présentés au tableau 7.6. Au total, 48 scénarii ($2 \times 1 \times 4 \times 6$) ont été analysés. Pour chaque scénario, nous avons simulé la production de 4 mois. Les deux premiers mois étant considérés comme période transitoire, les statistiques obtenues n'ont pas été prises en compte pour l'analyse des résultats afin d'éliminer des données biaisées dues à l'initialisation du modèle.

Tab. 7.6 – Facteurs et niveaux du plan d'expériences

Facteurs	Niveaux
Règle de sélection d'une tâche	STD, CL-FCFS
Règle de sélection d'un véhicule	LIV
Nombre de zones dans la fab	1, 5, 10, 15
Nombre de véhicules	18, 19, 20, 21, 24, 30

Les résultats des expériences de simulation obtenus pour le temps de cycle et le taux moyen d'utilisation des véhicules sont montrés par les graphiques de la figure 7.1. Pour des raisons de confidentialité, les valeurs chiffrées du temps de cycle ne peuvent être présentées. Afin de quantifier l'effet de la politique de gestion et le découpage en zones, ces valeurs sont remplacées par le X-factor qui est présenté aux tableaux 7.7, 7.8, 7.9 et 7.10 avec les autres résultats des simulations. Afin d'effectuer les comparaisons avec la gestion du système sans la mise en place du découpage du réseau de transport en zones, les premières colonnes des tableaux reprennent les résultats des simulations avec la politique traditionnelle de gestion des véhicules sans le découpage en zones (notée "1-zone").

Les tableaux 7.7 et 7.8 présentent les résultats obtenus avec la règle de dispatching STD (*Shostest Travel Distance*). En regardant ce tableau et la figure 7.1, nous pouvons observer une diminution du temps total de cycle pour certains scénarii, notamment avec une petite taille de la flotte de véhicules. Un résultat très intéressant concerne la diminution du taux

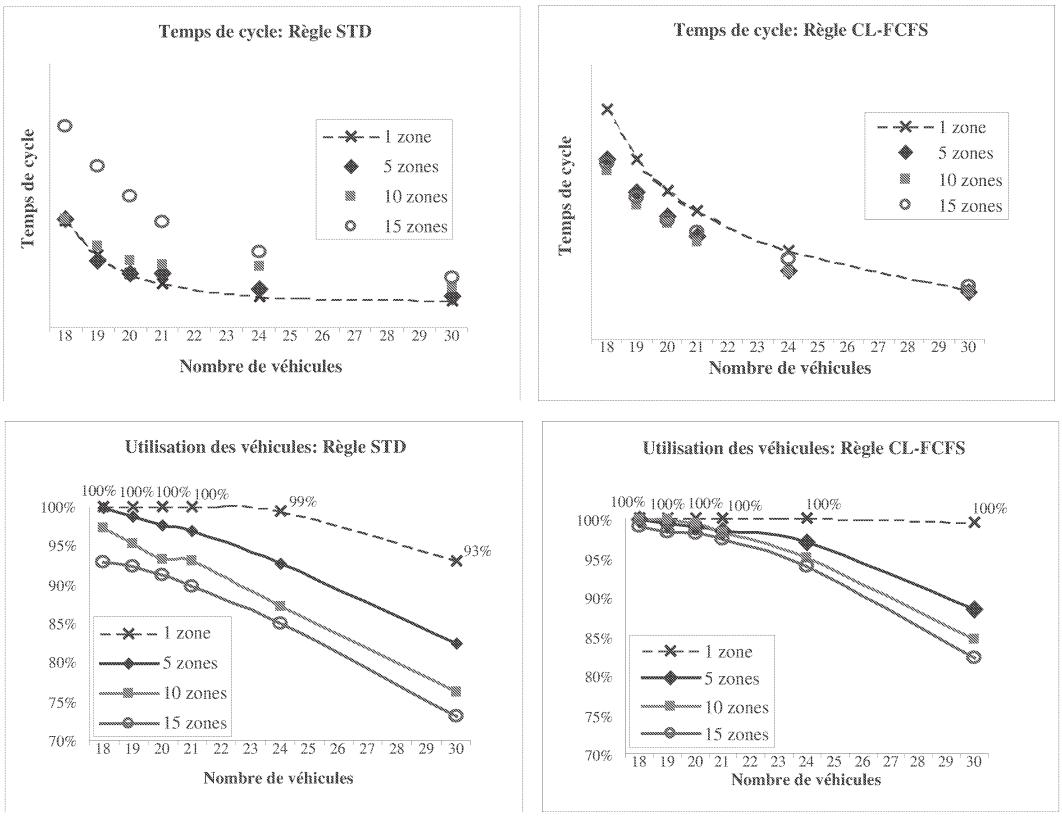


FIG. 7.1 – Comparaison des résultats : Temps de cycle et utilisation des véhicules

d'utilisation des véhicules par rapport à la gestion sans zones. Par exemple, avec 5 zones et 18, 19 et 20 véhicules, nous obtenons des temps de cycle équivalents et même inférieurs pour certains cas, tout en diminuant le taux d'utilisation des véhicules. Cela nous permet d'avoir plus de robustesse pour l'implémentation des politiques de maintenance réactive (suite aux pannes) ou préventive des véhicules. Avec une configuration de 10 zones et peu de véhicules dans le système, nous arrivons à des résultats similaires. Nous pouvons néanmoins observer des mauvaise valeurs pour le temps de cycle, par exemple avec une configuration avec 15 zones de gestion. Cela nous conduit à déduire que plus il y a des zones dans la fab, moins le système de transport est efficace lorsque la règle de dispatching STD est utilisée. Nous verrons par la suite que l'utilisation de la règle CL-FCFS permet d'améliorer le temps de cycle avec un grand nombre de zones.

Dans les tableaux 7.9 et 7.10, nous présentons les résultats obtenus avec la règle de dispatching CL-FCFS (*Current Location-First Come First Serve*). En regardant ce tableau et la figure 7.1, nous pouvons remarquer que le temps de cycle est toujours diminué lorsqu'une stratégie de gestion du transport par zones est mise en place. La diminution du taux d'utilisation des véhicules n'est pas aussi importante que nous l'avons observé pour la règle STD, notamment avec une petite taille de la flotte de véhicules. Lorsque le système a suffisamment de chariots (de l'ordre de 21 ou plus), leur niveau d'utilisation commence à diminuer, ainsi que le temps de cycle. Observons également qu'avec 30 véhicules, nous arrivons à saturation et nous ne pouvons plus diminuer le temps de cycle, alors que le

TAB. 7.7 – Résultats des simulations de l'implémentation de la planification tactique : Règle STD, 1 zone (scénario de référence) et 5 zones

	1-zone (référence)					
Nbr. véhicules	18	19	20	21	24	30
X-factor	1,56	1,44	1,38	1,35	1,31	1,30
Ecart-type du temps de cycle	0,17	0,09	0,07	0,06	0,04	0,04
Temps d'attente d'un véhicule (% du temps de non-production)	64%	58%	53%	50%	44%	43%
Utilisation maximale des machines	93%	93%	93%	93%	93%	93%
Utilisation de véhicules	100%	100%	100%	100%	99%	93%
Temps de livraison	1,47	1,54	1,62	1,70	1,94	2,27
Temps de calcul (hr.)	2,47	2,27	6,31	2,34	4,24	5,71
	5-zone					
Nbr. véhicules	18	19	20	21	24	30
X-factor	1,56	1,43	1,39	1,38	1,33	1,31
Ecart-type du temps de cycle	0,26	0,14	0,11	0,07	0,05	0,04
Temps d'attente d'un véhicule (% du temps de non-production)	60%	55%	51%	54%	49%	45%
Utilisation maximale des machines	93%	93%	93%	93%	93%	93%
Utilisation de véhicules	100%	99%	98%	97%	93%	82%
Temps de livraison	1,47	1,53	1,59	1,65	1,80	2,00
Temps de calcul (hr.)	28,8	17,4	21,5	59,9	0,5	5,0

TAB. 7.8 – Résultats des simulations de l'implémentation de la planification tactique : Règle STD, 10 et 15 zones

	10-zone					
Nbr. véhicules	18	19	20	21	24	30
X-factor	1,56	1,47	1,43	1,41	1,41	1,34
Ecart-type du temps de cycle	0,20	0,19	0,14	0,05	0,08	0,05
Temps d'attente d'un véhicule (% du temps de non-production)	61%	57%	54%	59%	55%	49%
Utilisation maximale des machines	93%	93%	93%	93%	93%	93%
Utilisation de véhicules	97%	95%	93%	93%	87%	76%
Temps de livraison	1,43	1,47	1,52	1,67	1,70	1,85
Temps de calcul (hr.)	115,3	82,0	98,9	31,4	53,7	6,0
	15-zone					
Nbr. véhicules	18	19	20	21	24	30
X-factor	1,87	1,74	1,64	1,56	1,46	1,37
Ecart-type du temps de cycle	1,89	1,17	0,37	0,17	0,10	0,06
Temps d'attente d'un véhicule (% du temps de non-production)	69%	67%	65%	63%	59%	52%
Utilisation maximale des machines	87%	91%	92%	93%	93%	93%
Utilisation de véhicules	93%	92%	91%	90%	85%	73%
Temps de livraison	1,50	1,50	1,50	1,54	1,66	1,78
Temps de calcul (hr.)	274,98	145,63	122,65	152,18	91,0	7,3

taux d'utilisation des véhicules continue à diminuer. Cela montre bien l'impact positif de la politique de gestion du transport par zones. Notamment, avec la règle de dispatching CL-FCFS, plus on diminue l'espace de recherche de tâches (i.e. plus on crée des zones), plus on améliore les indicateurs du système de transport.

TAB. 7.9 – Résultats des simulations de l'implémentation de la planification tactique : Règle CL-FCFS, 1 zone (scénario de référence) et 5 zones

	1-zone (référence)					
Nbr. véhicules	18	19	20	21	24	30
X-factor	1,94	1,78	1,68	1,61	1,48	1,36
Ecart-type du temps de cycle	0,16	0,08	0,07	0,06	0,05	0,04
Temps d'attente d'un véhicule (% du temps de non-production)	79%	77%	74%	72%	65%	53%
Utilisation maximale des machines	93%	93%	93%	93%	93%	93%
Utilisation de véhicules	100%	100%	100%	100%	100%	100%
Temps de livraison	1,46	1,54	1,62	1,70	1,95	2,42
Temps de calcul (hr.)	1,96	2,32	3,19	2,06	3,37	4,84
	5-zone					
Nbr. véhicules	18	19	20	21	24	30
X-factor	1,78	1,68	1,59	1,53	1,42	1,35
Ecart-type du temps de cycle	0,09	0,07	0,06	0,05	0,04	0,04
Temps d'attente d'un véhicule (% du temps de non-production)	76%	74%	71%	68%	60%	53%
Utilisation maximale des machines	93%	93%	93%	93%	93%	93%
Utilisation de véhicules	100%	99%	99%	98%	97%	88%
Temps de livraison	1,46	1,53	1,60	1,68	1,89	2,15
Temps de calcul (hr.)	25,1	47,7	68,4	87,8	0,6	9,5

TAB. 7.10 – Résultats des simulations de l'implémentation de la planification tactique : Règle CL-FCFS, 10 et 15 zones

	10-zone					
Nbr. véhicules	18	19	20	21	24	30
X-factor	1,74	1,63	1,57	1,52	1,42	1,36
Ecart-type du temps de cycle	0,08	0,06	0,06	0,05	0,05	0,04
Temps d'attente d'un véhicule (% du temps de non-production)	76%	72%	70%	67%	60%	53%
Utilisation maximale des machines	93%	93%	93%	93%	93%	93%
Utilisation de véhicules	100%	100%	99%	98%	95%	85%
Temps de livraison	1,46	1,54	1,61	1,67	1,85	2,05
Temps de calcul (hr.)	34,5	16,4	23,2	77,7	27,2	17,9
	15-zone					
Nbr. véhicules	18	19	20	21	24	30
X-factor	1,77	1,65	1,58	1,54	1,46	1,37
Ecart-type du temps de cycle	0,99	0,76	0,35	0,06	0,05	0,04
Temps d'attente d'un véhicule (% du temps de non-production)	77%	74%	70%	68%	63%	55%
Utilisation maximale des machines	92%	92%	93%	93%	93%	93%
Utilisation de véhicules	99%	98%	98%	97%	94%	82%
Temps de livraison	1,50	1,58	1,61	1,66	1,83	2,00
Temps de calcul (hr.)	30,4	23,8	16,5	144,5	196,1	23,1

Une remarque très importante concerne le temps de calcul nécessaire pour les simulations. Nous l'avons inclus aux tableaux 7.7, 7.8, 7.9 et 7.10 dans les dernières lignes de chaque scénario. Les temps de calcul sont donnés par le logiciel en secondes, mais nous les avons transformés en heures pour une meilleure lisibilité. Rappelons que les tests ont été réalisés

sur un PC Pentium® 4 (3,4 GHz). Selon le scénario testé, le temps nécessaire à la simulation peut atteindre 11 jours de calculs. Nous pensons que cela est dû au logiciel utilisé lui-même et non pas aux conditions matérielles ou de tests.

7.3.2 Etude sur le changement des paramètres tactiques

Dans cette section, nous allons analyser la cohérence entre les niveaux tactique et opérationnel de décision. Notre objectif est de constater qu'une mauvaise solution tactique donne aussi une mauvaise solution opérationnelle. Pour ce faire, nous allons procéder en deux étapes. Premièrement, nous allons modifier la distribution des véhicules dans les zones tout en gardant leur configuration (i.e. définition des zones). Les solutions de simulations seront comparées avec les cas où il n'y a pas eu de modifications (solutions optimales de la planification tactique). Deuxièmement, nous allons considérer de mauvaises solutions pour le temps de couverture. La configuration des zones et l'affectation des véhicules obtenues seront implémentées sur le modèle de simulation, et les performances opérationnelles seront comparées.

7.3.2.1 Changement du nombre des véhicules affectés aux zones

Dans cette section, nous allons réaliser une première série d'expériences dans le but de vérifier la cohérence entre les niveaux de planification tactique et de gestion opérationnelle. Pour ce faire, nous allons modifier la distribution des véhicules dans les zones tout en gardant leur configuration (i.e. choix des points).
Dans cette étude, nous allons découper le réseau de transport en 5 zones et un total de 18, 20 et 24 véhicules devront être affectés à chacune des zones. Le tableau 7.11 présente les distributions des véhicules. Pour chaque flotte, la colonne "NT" indique la distribution obtenue à partir de l'optimisation tactique, et la colonne "Mod." indique la modification arbitraire que nous avons effectuée. Rappelons que pour chaque scénario testé, le découpage du réseau en zones (i.e. solution tactique) peut changer pour chaque couple <Nombre de véhicules-Nombre de zones> : la configuration avec 5 zones et 18 véhicules n'est pas forcément la même qu'avec 20 ou 24 véhicules, et ainsi de suite.

Tab. 7.11 – Distribution des véhicules dans les zones

Scénario*	NT	Mod.		NT	Mod.		NT	Mod.
Zone 1(18v)	3	4	Zone 1(20v)	8	3	Zone 1(24v)	5	9
Zone 2(18v)	7	2	Zone 2(20v)	2	5	Zone 2(24v)	9	3
Zone 3(18v)	2	2	Zone 3(20v)	2	6	Zone 3(24v)	3	5
Zone 4(18v)	3	5	Zone 4(20v)	3	4	Zone 4(24v)	4	2
Zone 5(18v)	3	5	Zone 5(20v)	5	2	Zone 5(24v)	3	5
Totaux	18	18		20	20		24	24

*NT : solution optimale du niveau tactique
Mod. : affectation modifiée

Le tableau 7.12 présente les résultats obtenus pour les deux règles de dispatching STD et CL-FCFS. Nous observons bien une déterioration des principaux indicateurs de performance. Le temps de cycle augmente de 4% ou 8% avec STD et de 3% ou 6% avec CL-FCFS. Le temps d'attente pour l'arrivée d'un chariot peut augmenter jusqu'à 55% avec STD et

jusqu'à 20% avec CL-FCFS, pour la configuration avec 20 véhicules. Avec 18 véhicules les écarts sont moins importants simplement parce qu'on se rapproche de la taille minimale requise pour servir les demandes (rappelons qu'en utilisant l'équation (6.11) présentée au chapitre 6, la taille minimale de la flotte pour avoir une solution faisable pour les programmes linéaires est de 18 véhicules). Les diminution observées pour les taux d'utilisation des véhicules sont expliquées par les "mauvaises" distributions de véhicules (i.e. affectation aux zones) : la mauvaise distribution des véhicules dans les zones fait qu'à certains moments des véhicules sont disponibles sans pouvoir satisfaire les demandes des autres zones (la politique opérationnelle ne le leur permet pas).

TAB. 7.12 – Résultats de la modification de la distribution des véhicules dans les zones
Nombre de zones : 5

Règle de dispatching : STD						
Nbr. de véhicules	18		20		24	
Scénario*	NT	Mod.	NT	Mod.	NT	Mod.
X-factor	1,56	1,62	1,39	1,49	1,33	1,39
Ecart-type du temps de cycle	0,26	0,92	0,11	0,45	0,05	0,24
Temps d'attente d'un véhicule (% du temps de non-production)	60%	68%	51%	62%	49%	55%
Utilisation maximale des machines	93%	91%	93%	92%	93%	93%
Utilisation des véhicules	100%	96%	98%	92%	93%	86%
Temps de livraison	1,47	1,46	1,59	1,52	1,80	1,70
Variation du temps de cycle		+4%		+8%		+4%
Variation du temps d'attente d'un véhicule		+26%		+55%		+32%
Temps de calcul (hr.)	28,8	28,3	21,5	24,1	0,5	25,6
Règle de dispatching : CL-FCFS						
Nbr. de véhicules	18		20		24	
Scénario*	NT	Mod.	NT	Mod.	NT	Mod.
X-factor	1,78	1,83	1,59	1,68	1,42	1,46
Ecart-type du temps de cycle	0,09	0,10	0,06	0,07	0,04	0,05
Temps d'attente d'un véhicule (% du temps de non-production)	76%	77%	71%	74%	60%	64%
Utilisation maximale des machines	93%	93%	93%	93%	93%	93%
Utilisation des véhicules	100%	99%	99%	95%	97%	92%
Temps de livraison	1,46	1,45	1,60	1,54	1,89	1,80
Variation du temps de cycle		+3%		+6%		+3%
Variation du temps d'attente d'un véhicule		+8%		+20%		+16%
Temps de calcul (hr.)	25,1	23,7	68,4	105,4	0,6	143,3

*NT : solution optimale du niveau tactique
Mod. : distribution modifiée des véhicules

7.3.2.2 Ajout d'une borne pour le temps de couverture

Nous allons maintenant forcer le modèle tactique à donner des mauvaises solutions pour le temps de couverture. Pour ce faire, nous allons introduire une nouvelle contrainte dans notre formulation mathématique du problème de définition des zones et de positionnement des véhicules. Cette nouvelle contrainte est présentée ci-dessous :

$$T \geq TB \tag{7.1}$$

Où TB est une valeur nettement supérieure à la valeur optimale du temps maximum de couverture du niveau tactique. Par exemple, nous avons vu au chapitre 6 que la valeur de T lorsque nous voulons créer 5 zones était de 73. Dans notre étude, nous allons prendre $TB = \{80, 100, 120\}$ et résoudre le modèle avec une flotte de 24 véhicules. Les solutions obtenues à partir de la résolution des modèles tactiques (i.e. la configuration des zones et l'affectation des véhicules) sont implémentées sur le modèle de simulation, et les performances opérationnelles comparées.

Le tableau 7.13 présente les résultats obtenus pour les deux règles de dispatching STD et CL-FCFS. La première colonne du tableau reprend la solution obtenue avant l'ajout de la contrainte (7.1). Le fait de forcer le modèle tactique à donner une mauvaise solution de découpage, provoque un impact négatif sur la performance opérationnelle. Selon les conditions de test choisies, le temps de cycle est en effet déterioré d'entre 2% et 4% avec la règle de dispatching STD, et d'entre 3% et 6% avec la règle CL-FCFS. Cette augmentation du temps de cycle est notamment due à l'augmentation du temps d'attente d'un chariot par l'entité qui peut atteindre +28% avec STD ou encore +34% avec CL-FCFS.

TAB. 7.13 – Résultats des simulations avec une borne pour le temps de couveture

Nombre de zones : 5
Nombre de véhicules : 24

Règle de dispatching : STD				
Valeur de la borne TB	sans (7.1)	80	100	120
X-factor	1,33	1,36	1,37	1,38
Ecart-type du temps de cycle	0,05	0,06	0,07	0,06
Temps d'attente d'un véhicule (% du temps de non-production)	49%	52%	53%	54%
Utilisation maximale des machines	93%	93%	93%	93%
Utilisation des véhicules	93%	92%	94%	91%
Temps de livraison	1,80	1,79	1,83	1,76
Variation du temps de cycle		+2%	+3%	+4%
Variation du temps d'attente d'un véhicule		+18%	+23%	+28%
Temps de calcul (hr.)	0,5	57,0	31,0	19,8
Règle de dispatching : CL-FCFS				
Valeur de la borne TB	sans (7.1)	80	100	120
X-factor	1,42	1,47	1,50	1,51
Ecart-type du temps de cycle	0,04	0,22	0,05	0,05
Temps d'attente d'un véhicule (% du temps de non-production)	60%	64%	66%	67%
Utilisation maximale des machines	93%	93%	93%	93%
Utilisation des véhicules	97%	97%	98%	97%
Temps de livraison	1,89	1,91	1,91	1,88
Variation du temps de cycle		+3%	+5%	+6%
Variation du temps d'attente d'un véhicule		+18%	+30%	+34%
Temps de calcul (hr.)	0,6	8,7	73,9	44,0

7.4 Sur l'impact de véhicules partagés

Jusqu'à présent, nous avons considéré une flotte de véhicules donnée où tous les chariots sont affectés aux zones créées pour satisfaire les demandes de transport. Comme nous avons expliqué dans la section 7.1, nous pourrions être intéressés par une analyse de la performance du système lorsque certains véhicules sont affectés aux zones alors que d'autre ne le sont pas.
Dans cette section, nous allons donc introduire le concept de "véhicules partagés" et analyser son impact sur la performance opérationnelle d'une gestion par zones du transport dans la fab. Avant d'effectuer cette analyse, nous avons besoin de définir ce qu'est un véhicule "partagé". Lorsque le transport des lots dans la fab est effectué par zones de gestion comme nous les avons définies précédemment, un véhicule est dit "partagé" s'il n'est pas attaché à une zone particulière et par conséquent, peut servir n'importe quelle demande dans le système, selon la ou les règles de dispatching qui lui ont été définies. Remarquons que si tous les véhicules sont "partagés", cela équivaut à ne pas avoir de zones.

L'introduction de véhicules partagés dans le réseau permettrait une meilleure flexibilité pour servir les requêtes en attente d'un chariot. Lorsqu'une flotte de véhicules est affectée à une zone particulière, ceux-ci ne peuvent servir que les demandes provenant de cette zone. Il se peut donc qu'un instant donné, une zone ait des véhicules libres (car ils sont affectés mais il n'y a pas de demandes de transport) alors que dans une autre zone les demandes s'accumulent et ne sont pas satisfaites car ses véhicules sont tous occupés. Nous pourrions améliorer les performances opérationnelles du système de transport si une stratégie complémentaire, comme par exemple les véhicules partagés, était mise en place. Cette stratégie permettra de réduire les temps d'attente des demandes ne provenant pas de zones où des véhicules sont disponibles et même ajouter de la robustesse pour la gestion globale des véhicules, donnant ainsi plus de flexibilité pour l'exécution des activités de maintenance préventive ou réactive du système de transport.

Afin d'évaluer l'impact d'une flotte de véhicules partagés, nous avons effectué une série d'expériences de simulation avec la politique opérationnelle ZVBDemFab. Rappelons qu'avec cette politique, la zone dont le point j est le centre se voit affecter une flotte de $\sum_v Z_{vj}$ véhicules. De plus, nous définissons un certain nombre de véhicules qui resteront "non affectés". Comme expliqué dans le paragraphe précédent, ces véhicules dits "partagés" peuvent servir n'importe quelle demande dans le système sans se préoccuper de sa provenance.

Lors des tests de simulations, nous avons découpé le réseau en 5 zones de gestion. Le nombre total de véhicules est fixé à 24. Le nombre de véhicules partagés varie entre 0 et 24. Le tableau 7.14 présente les résultats obtenus. Avec une flotte de 24 véhicules et la mise en œuvre de la règle STD pour le dispatching il est possible de réduire le temps de cycle jusqu'à 2,1% alors que la réduction du temps d'attente d'un véhicule par l'entité peut attendre 16,5%. Avec 24 véhicules et la règle CL-FCFS, les résultats sont moins encourageants. Les réductions du temps de cycle et du temps d'attente d'attente effective d'un chariot avec la règle de dispatching STD sont très positives, respectivement entre 0,3% et 2,1% et entre 3,7% et 16,5%. Ces résultats nous permettent de constater aussi la cohérence entre la planification tactique et la gestion opérationnelle. Avec la règle de dispatching CL-FCFS,

les résultats sont moins encourageants, le temps de cycle pouvant augmenter jusqu'à 4,4%, et le temps d'attente d'un véhicule par l'entité peut lui-aussi augmenter jusqu'à 24%. Les mauvaises performances de cette implementantion peuvent être dues aux spécificités de la politique opérationnelle mise en œuvre.

TAB. 7.14 – Résultats des simulations : Politique ZVBDemFab avec véhicules partagés

Nombre de zones : 5					
Règle de dispatching : STD					
Nombre total de véhicules	24	24	24	24	24
(dont véhicules "partagés")	0	6	12	18	24
Pourcentage de partage de la flotte	0%	25%	50%	75%	100%
X-factor	1,33	1,35	1,33	1,31	1,31
Ecart-type du temps de cycle	0,05	0,06	0,06	0,05	0,04
Temps d'attente d'un véhicule (% du temps de non-production)	49%	50%	47%	45%	44%
Utilisation maximale des machines	93%	93%	93%	93%	93%
Utilisation des véhicules affectés	93%	87%	85%	83%	-
Utilisation des véhicules partagés	-	100%	100%	100%	99%
Utilisation moyenne des véhicules	93%	93%	93%	91%	99%
Temps moyen de livraison	1,80	1,73	1,81	1,89	1,94
Variation du temps de cycle		1,5%	-0,3%	-1,4%	-2,1%
Variation du temps d'attente d'un véhicule		10,2%	-3,7%	-11,9%	-16,5%
Règle de dispatching : CL-FCFS					
Nombre total de véhicules	24	24	24	24	24
(dont véhicules "partagés")	0	6	12	18	24
Pourcentage de partage de la flotte	0%	25%	50%	75%	100%
X-factor	1,42	1,43	1,44	1,46	1,48
Ecart-type du temps de cycle	0,04	0,04	0,05	0,05	0,05
Temps d'attente d'un véhicule (% du temps de non-production)	60%	61%	62%	64%	65%
Utilisation maximale des machines	93%	93%	93%	93%	93%
Utilisation des véhicules affectés	97%	96%	98%	99%	-
Utilisation des véhicules partagés	-	100%	100%	100%	100%
Utilisation moyenne des véhicules	97%	98%	99%	99%	100%
Temps moyen de livraison	1,89	1,95	1,93	1,89	1,95
Variation du temps de cycle		0,4%	1,2%	2,9%	4,4%
Variation du temps d'attente d'un véhicule		2,1%	6,6%	15,7%	24,3%

7.5 Conclusion

Dans ce chapitre, nous avons étudié la problématique de gestion opérationnelle de véhicules dans un réseau découpé en zones. Nous sommes partis des résultats obtenus par l'optimisation du niveau de planification tactique pour les incorporer dans le modèle de simulation de la fab. Afin de prendre en compte l'évolution dynamique du système, plusieurs politiques opérationnelles de gestion des véhicules peuvent être mises en place. Nous avons analysé certaines d'entre elles, qui se différencient particulièrement de la stratégie utilisée pour la gestion des demandes de transport et des véhicules au cours de la période de fabrication. Nous avons aussi réalisé une analyse plus étendue avec la politique de gestion donnant les meilleures performances (appelée ZVBDemFab), et l'avons comparée avec les résultats d'une gestion du transport sans zones. Cette étude nous a montré l'intérêt de mettre en

place la gestion par zones, notamment pour la réduction du temps de cycle et/ou pour incorporer de la flexibilité pour la gestion des véhicules, selon la règle de dispatching utilisée. Nous avons également effectué une étude sur la cohérence entre les niveaux tactique et opérationnel de décision. Cette étude a été réalisée en deux étapes. Premièrement, nous avons arbitrairement changé la distribution des véhicules dans les zones. Nous avons fait varier la taille de la flotte afin d'analyser les performances du système au niveau opérationnel. Deuxièmement, nous avons forcé le modèle d'optimisation tactique à donner une mauvaise solution pour le temps maximal de couverture, en ajoutant une contrainte additionnelle aux formulations mathématiques. Cette solution a été mise en place au niveau opérationnel. Les performances obtenues avec ces deux modifications du modèle de simulation ont montré la cohérence entre les décisions de la planification tactique et le comportement opérationnel du système. A la fin du chapitre, nous avons testé l'intérêt de l'implémentation d'une stratégie complémentaire à la politique opérationnelle qui permet d'avoir une flotte de véhicules non affectés aux zones créées lors de la planification tactique. Cette stratégie permet donc d'avoir un certain nombre de véhicules "partagés" capables de servir n'importe quelle demande de transport sans se préocuper de sa provenance dans le réseau. Nous avons constaté l'impact positif aussi bien en termes de réduction du temps de cycle que pour les autres indicateurs de performance du système de transport.

Les résultats expérimentaux obtenus en simulation nous montrent l'intérêt de l'implémentation d'une gestion du transport de produits par zones, en couplant des techniques d'optimisation pour la planification tactique et des modèles de simulation pour l'analyse dynamique du comportement opérationnel du système. Nos travaux peuvent bien sûr donner lieu à plusieurs extensions, notamment lors de la mise en œuvre industrielle. Certaines considérations pratiques doivent être prises en compte. Nous les analysons plus amplement dans les conclusions et perspectives de cette thèse.

Chapitre 8

Conclusion générale et perspectives

Dans cette thèse, nous nous sommes intéressés au problème de gestion du transport dans les unités fabrication automatisée de semi-conducteurs (fabs). Nous avons d'abord commencé par une présentation du contexte industriel et de la conjoncture économique actuelle afin de souligner l'importance de l'automatisation du transport des encours dans les fabs. Ensuite, nous avons présenté une description générale du processus de fabrication de wafers. L'objectif principal de cette description était d'appréhender et de mieux comprendre les caractéristiques et la complexité intrinsèque du processus de fabrication de wafers. Ce chapitre a été suivi d'une étude de la littérature traitant de la conception et de la gestion du système de transport automatisé dans les fabs. Nous avons identifié les manques de la littérature scientifique sur le pilotage de tels systèmes. A partir de nos conclusions, nous avons établi notre démarche de recherche. Nous avons ainsi défini quatre étapes principales : modélisation conceptuelle, simulation et analyse du système, optimisation, et couplage optimisation-simulation.

Premièrement, nous avons utilisé des méthodologies de modélisation pour identifier les composantes de la fab (e.g. système de supervision, machines, véhicules, etc.) et formaliser leurs relations (flux physiques et d'information). Notre objectif était de décrire, à l'aide de formalismes de modélisation, le système de production et le processus de fabrication, en intégrant les aspects production et transport. Ce cadre de modélisation permet également d'identifier les entités du système et leurs interactions afin de mettre en place une démarche de simulation et d'optimisation. Ces modèles conceptuels nous ont servi ensuite à mieux comprendre, d'une part, où sont situés nos travaux à l'intérieur du système global de supervision de la fab, et d'autre part, à construire le modèle de simulation à événements discrets utilisé pour l'analyse dynamique du système. Selon la terminologie définie dans [42], cette partie de notre travail constitue la *spécification* du modèle lors d'une analyse par simulation. Notre modèle conceptuel nous a donc permis de savoir ce que notre système *fait*.
A partir de notre modèle conceptuel, nous avons entamé la deuxième étape de notre travail. Après une présentation détaillée de notre étude de cas (une nouvelle fab dédiée à la fabrication de wafers de 300mm), nous avons réalisé des expériences de simulation dans le but d'identifier les paramètres critiques et pertinents pour l'optimisation de la performance industrielle et celle du système de transport. En reprenant les définitions présentées dans [42], cette approche constitue l'*implémentation* du modèle conceptuel, avec laquelle nous avons pu voir *comment* se comporte notre système. Lors de notre analyse par simulation, nous avons aussi pu tester et comparer plusieurs règles de dispatching des véhicules

et de gestion du réseau de transport. Les résultats obtenus étant très encourageants, nous avons proposé une démarche d'optimisation. Dans cette troisième étape, l'objectif était d'améliorer les principaux indicateurs de performance de la fab. Notre démarche d'optimisation a été réalisée sur deux niveaux : tactique et opérationnel.

Le niveau tactique a consisté à étudier le problème du point de vue global afin de décider de la "meilleure" façon de gérer les véhicules dans le système. Notre objectif était donc de définir la distribution des véhicules qui permettra de diminuer au maximum le temps de service d'une demande de transport. Celui-ci était l'objectif principal. Cependant, dans l'intérêt d'améliorer les performances du système de transport, nous avons aussi pris en compte les taux d'utilisation des véhicules et les demandes de services dans le système. Cette approche nous a conduit à proposer un modèle d'optimisation multi-objectif. La solution du niveau tactique nous permet d'obtenir un découpage du réseau de transport en zones de gestion et d'affecter une flotte de véhicules à chacune des zones permettant de satisfaire les requêtes de transport. Certains compromis peuvent être faits entre les variables du modèle, lesquels doivent être vérifiés, au niveau opérationnel, par une étude de simulation.
Ainsi, la dernière étape de notre démarche d'analyse et gestion du système de transport a consisté à traiter le problème du point de vue opérationnel en couplant la solution optimale du niveau tactique avec un modèle permettant de prendre en compte l'évolution dynamique du système de fabrication. Pour ce faire, nous avons dû définir des politiques pour la gestion opérationnelle des véhicules. Nous avons donc défini et comparé plusieurs stratégies, chacune caractérisée par des critères dynamiques pour la ré-affectation des véhicules. Nos résultats ont également montré que cette gestion par zones est bien plus intéressante lorsqu'une stratégie appropriée de dispatching de véhicules est implémentée. En effet, nous avons pu voir que la règle de dispatching FCFS (*First Come First Serve*) ou FIFO (*First In First Out*) qui a été le plus utilisée dans la littérature, donne de très mauvaises performances si le réseau n'est pas découpé en zones. Avec le découpage en zones et une politique intelligente de gestion des véhicules, la règle FCFS peut être très intéressante. La politique de dispatching STD (*Shortest Travel Distance*) a aussi été implémentée, et les performances du système analysées avec et sans mise en place de la gestion du transport par zones. Avec la règle STD, le découpage en zones n'a pas toujours réduit le temps de cycle dans les cas testés. Néanmoins, l'analyse de la performance industrielle étant multi-critère, nous avons obtenu une valeur du temps de cycle équivalent (par rapport à une gestion sans zones) avec en plus une réduction de la charge de travail subie par les véhicules (i.e. diminution de leur taux d'utilisation). Ceci permet donc d'avoir plus de robustesse pour la gestion globale du système de transport, notamment pour la mise en œuvre de politiques de maintenance préventive ou réactive.

La méthode proposée est appliquée à un cas industriel du domaine de la microélectronique. Les tests réalisés par simulation montrent que les hypothèses de l'optimisation sont valables et que celle-ci est applicable rapidement puisque les objectifs visés tiennent compte des aspects opérationnels de l'unité de fabrication.

Pour faire suite à ce travail, nous envisageons trois principales perspectives. La première d'entre elles concerne la proposition de méthodes heuristiques ou méta-heuristiques efficaces pour résoudre la problématique du niveau tactique. Cette ligne de travail s'inscrit

aussi dans une démarche dynamique d'aide à la décision. Nous avons commencé avec la proposition d'une méthode simple basée sur le principe glouton. L'idée maintenant est de proposer des méthodes méta-heuristiques plus intelligentes et plus rapides et qui pourraient prendre en compte des relations non-linéaires entre les variables.

Une deuxième perspective de cette thèse concerne l'application de l'approche tactique proposée pour déterminer la localisation des machines dans la fab. Dans notre étude, nous avons considéré l'agencement des machines comme étant fixé. Nous pourrions relaxer cette hypothèse et utiliser les résultats du niveau tactique (i.e. la définition de zones) et quelques simulations pour évaluer la pertinence de la méthodologie actuellement employée par les fabricants pour la définition du *layout*. En effet, les modèles utilisés pour la résolution du problème tactique de définition de zones se basent sur des informations obtenues à partir de la demande de chaque point du réseau. Cette demande équivaut en fait au nombre de lots sortant de chaque point (machine). Nous pouvons utiliser le regroupement par zones pour identifier les machines qui devraient être localisées dans une même zone afin de diminuer le temps de transfert des lots entre les opérations de production. Des contraintes technologiques devront bien sûr être respectées (e.g. isolement des machines de photolithographie). La figure 8.1 ci-après présente la façon dont nos modèles d'optimisation et de simulation peuvent s'intégrer dans une démarche générale, à la fois pour l'optimisation de la gestion globale du système et pour l'aide à la décision. Cette figure montre bien le rôle qui jouent les modèles de simulation et d'optimisation.

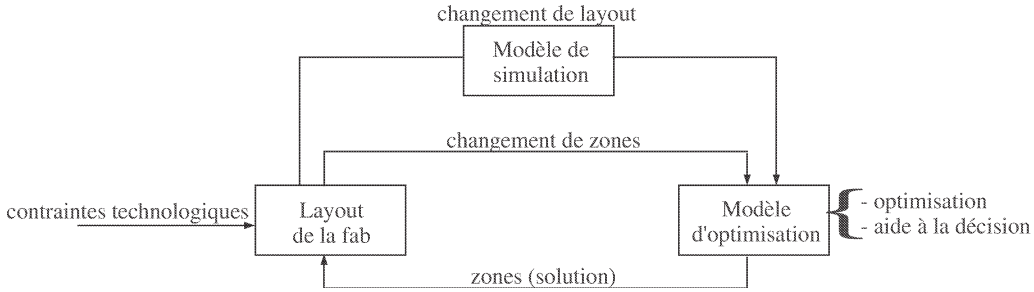


Fig. 8.1 – Intégration de notre démarche pour l'optimisation, la simulation et l'aide à la décision

Enfin, la troisième perspective de travail qui est très intéressante du point de vue de la gestion globale de la fab, concerne l'étude de l'intégration de stratégies de contrôle de véhicules avec les politiques d'ordonnancement de la production et de lancement des lots. Dans notre recherche, nous avons pris en compte certaines contraintes liées à la production (e.g. stockage intermédiaire, files d'attente limitées devant les machines, production à flux tiré), mais nous n'avons pas testé l'impact des politiques d'ordonnancement. Plusieurs auteurs ont précédemment analysé les bénéfices de la mise en place de diverses politiques de contrôle de production [123, 53]. Avec l'automatisation du transport de wafers, il s'agirait maintenant d'intégrer ces deux aspects et de tester leur impact sur la performance globale de la fab. D'un point de vue théorique, les chercheurs se sont déjà intéressés aux problématiques de l'intégration de la gestion de la production et du transport (e.g. [117, 54, 7, 63, 62]). Du point de vue pratique, quelques travaux ont aussi été proposés dans la littérature pour d'autres types de systèmes de production et donnent des résultats

intéressants (e.g. [116, 24]). A notre connaissance, des analyses formelles de l'intégration entre production et transport et de l'impact de leurs interactions n'ont pas été réalisées dans les systèmes de fabrication de semi-conducteurs. Cette étude est un véritable challenge et une ligne de travail très intéressante. Dans un premier temps, le travail peut se focaliser sur l'intégration de ces dimensions de la problématique opérationnelle. Ensuite, le but serait d'étudier des politiques proactives permettant d'anticiper les besoins en produits des machines. Cette dernière est d'autant plus importante lorsque des lots prioritaires sont pris en compte.

Bibliographie

[1] Albert, P. *Pilotage de véhicules autoguidés dans un système automatisé de production.* PhD thesis, Université de Nantes, 1998.

[2] Ansquer, R. Modélisation d'équipements de type "cluster" et application à l'étude des rendements. Projet de fin d'études, IUP Génie de Matériaux, Université d'Evry Val d'Essonne, France, 1996.

[3] Arends, N. *A system engineering specification formalism.* PhD thesis, Technische Universiteit Eindhoven, The Netherlands, 1996.

[4] Askin, R., and Goldberg, J. *Design and analysis of lean production systems.* John Wiley & Sons, Inc., 2002.

[5] Aurand, S., and Miller, P. The operating curve : a method to measure and benchmark manufacturing line productivity. In *Proceedings of the 1997 IEEE/SEMI Advanced Semiconductor Manufacturing Conference* (1997), pp. 165–168.

[6] Barton, R. Designing simulation experiments. In *Proceedings of the 2002 Winter Simulation Conference* (2002), pp. 45–51.

[7] Bécart, M., Lacomme, P., Moukrim, A., and Tchernev, N. Proposition de résolution exacte d'un milp pour les systèmes flexibles de production du type job-shop avec transport. In *Actes de la 5e Conférence Francophone de Modélisation et Simulation "Modélisation et Simulation pour l'Analyse et l'Optimisation des Systèmes Industriels et Logistiques"* (2004), vol. 2, pp. 635–642.

[8] Beamon, B., and Chen, V. Performability-based fleet sizing in a material handling system. *International Journal of Advanced Manufacturing Technology 14*, 6 (1998), 441–449.

[9] Beamon, B., and Deshpande, A. Performance-based unit-load optimization in material handling system design. *Production Planning & Control 9*, 7 (1998), 634–639.

[10] Becker, M. Modeling and simulation of a complete semiconductor manufacturing facility using petri nets. In *Proceedings of the IEEE International Conference on Emerging Technologies and Factory Automation* (2003), vol. 2, pp. 153–156.

[11] Bertalanffy, L. V. *General System Theory.* George Braziller Inc., 1968.

[12] Brotcorne, L., Laporte, G., and Semet, F. Ambulance location and relocation models. *European Journal of Operational Research 147* (2003), 451–463.

[13] Burkard, R. Quadratic assignment problems. *European Journal of Operational Research 15* (1984), 283–289.

[14] Cardarelli, G., and Pelagagge. Simulation tool for design and management optimization of automated interbay material handling and storage systems for large wafer fab. *IEEE Transactions on Semiconductor Manufacturing 8*, 1 (1995), 44–49.

[15] Carpenter, B., Gibson, R., and Pierce, N. Generic interbay automated material handling system discrete-event simulation. Tech. Rep. 93011441A-GEN, International SEMATECH, 1993.

[16] Castagna, P., Mebarki, N., and Maza, S. Evaluation du nombre d'agvs sur un système de production de type job-shop. In *Actes de la 5e Conférence Francophone de Modélisation et Simulation "Modélisation et Simulation pour l'Analyse et l'Optimisation des Systèmes Industriels et Logistiques"*, vol. 2, pp. 619–626.

[17] Chang, C., and Chang, S. A layer-based layout approach for semiconductor fabrication facilities. In *Proceedings of the 1998 IEEE/SEMI Advanced Semiconductor Manufacturing Conference and Workshop* (1998), pp. 385–390.

[18] Chapron, J. *Pilotage de la transformation des processus informationnels de l'entreprise : Etude de cas de la microélectronique.* PhD thesis, Ecole Nationale Supérieure des Mines de Saint-Etienne, 2005. Unpublished.

[19] Chen, J., Fu, L., Lin, M., and Huang, A. Petri-net and ga-based approach to modeling, scheduling, and performance evaluation for wafer fabrication. *IEEE Transactions on Robotics and Automation 17*, 5 (2001), 619–636.

[20] Co, C., and Tanchoco, J. A review of research on agvs vehicle management. *Engineering Costs and Production Economics 21* (1991), 35–42.

[21] Dash Optimization. *Xpress MP manual*, 2004.

[22] Daskin, M. *Network and discrete location : Models, algorithms and applications.* John Wiley & Sons, Inc., 1995.

[23] DeJong, C., and Wu, S. Simulating the transport and scheduling of priority lots in semiconductor factories. In *Proceedings of the 2002 Winter Simulation Conference* (2002), pp. 1387–1391.

[24] Deroussi, L., Gourgand, M., and Tchernev, N. Coupling local search methods and simulated annealing to the job shop scheduling problem with transportation. In *Proceedings of the 8th IEEE International Conference on Emerging Technologies and Factory Automation* (2001), vol. 1, pp. 659–667.

[25] Duenyas, L., Fowler, J., and Schruben, L. Planning and scheduling in japanese semiconductor manufacturing. *Journal of Manufacturing Systems 13*, 5 (1994), 323–332.

[26] Durand, D. *La systémique "Que sais-je ?"*. PUF, 1979.

[27] Egbelu, P. Positioning of automated guided vehicles in a loop layout to improve response time. *European Journal of Operational Research 71* (1993), 32–44.

[28] Egbelu, P., and Tanchoco, J. Characterization of automatic guided vehicle dispatching rules. *International Journal of Production Research 22*, 3 (1984), 359–374.

[29] Fu, M., and Kaku, B. Minimizing work-in-process and material handling in the facilities layout problem. Tech. Rep. TR-95-41, Institute for Research Systems, University of Maryland, USA, 1995.

[30] Gademann, A., and van de Velde, S. Positioning automated guided vehicles in a loop layout. *European Journal of Operational Research 127* (2000), 565–573.

[31] Ganesharajah, T., Hall, N., and Sriskandarajah, C. Design and operational issues in agv-served manufacturing systems. *International Journal of Production Research 76* (1998), 109–154.

[32] Geiger, C., Hase, R., Takoudis, C., and Uszoy, R. Alternative layouts for semiconductor wafer fabrication facilities. *IEEE Transactions on Components, Packaging, and Manufacturing Technology 20*, 2 (1997), 152–163.

[33] Glassey, C., and Resende, M. Closed-loop job release control for vlsi circuit manufacturing. *IEEE Transactions on Semiconductor Manufacturing 1*, 1 (1988), 36–46.

[34] Grimaud, F. *Une démarche d'aide à la décision pour la conception et la gestion des systèmes industriels.* Lecture Notes. Ecole Nationale Supérieure des Mines de Saint-Etienne, 2002.

[35] Groover, M. *Fundamentals of modern manufacturing : materials, processes, and systems.* Prentice Hall, Englewood Cliff, 1996.

[36] Guo, R., Slama, M., Griffin, R., and Holman, K. A work cell manufacturing system for vlsi fabrication. In *Proceedings of the 1993 IEEE/CHMT International Electronics Manufacturing Technology Symposium* (1993), pp. 200–205.

[37] Haagh, P., Wilkens, A., van Campen, E., Rooda, J., and Rulkens, H. Application of a layout design method to a dielectrics deposition area in a 300mm wafer fab. In *Proceedings of the Seventh International Symposium on Semiconductor Manufacturing* (1998), pp. 69–72.

[38] Hase, R., Takoudis, C., and Uszoy, R. Cellular and reentrant layouts for semiconductor wafer fabrication facilities. In *Proceedings of the 1994 IEEE/CPMT International Electronics Manufacturing Technology Symposium* (1994), pp. 112–118.

[39] Hetzel, W. Cycle time reduction and strategic inventory placement across a multistage process, 1993.

[40] Hogg, G., Fowler, J., and Ibrahim, M. Flow control in semiconductor manufacturing : A survey and projection of needs. Tech. Rep. 91110757A-GEN, International SEMATECH, 1991.

[41] ITRS. International technology roadmap for semiconductors, annual report, 2003. http ://public.itrs.net.

[42] Jacobson, S., and Yücesan, E. On the complexity of verifying structural properties of discrete event simulation models. *Operations Research 47*, 3 (1999), 476–481.

[43] Jacobson, S., and Yücesan, E. Common issues in discrete optimization and discrete-event simulation. *IEEE Transactions on Automatic Control 47*, 2 (2001), 321–345.

[44] Jang, J., Suh, J., and Ferreira, P. An agv routing policy reflecting the current and future state of semiconductor and lcd production lines. *International Journal of Production Research 39*, 17 (2001), 3901–3921.

[45] Jeng, M., Xie, X., and Chou, S. Modeling, qualitative analysis, and performance evaluation of the etching area in an ic wafer fabrication system using petri nets. *IEEE Transactions on Semiconductor Manufacturing 11* (1998), 358–373.

[46] Jimenez, J., Kim, B., Fowler, J., Mackulak, G., Chung, Y., and Kim, D. Operational modeling and simulation of an inter-bay amhs in semiconductor wafer fabrication. In *Proceedings of the 2002 Winter Simulation Conference* (2002), pp. 1377–1382.

[47] Johri, P. Practical issues in scheduling and dispatching in semiconductor wafer fabrication. *Journal of Manufacturing Systems 12*, 6 (1993), 474–485.

[48] Kellert, P., and Ruch, S. Méthodologie de modélisation orientée objets de systèmes de production - un processus de construction/validation du modèle générique orienté objets d'un système de production. *Journal Européen des Systèmes Automatisés 32*, 1 (1998), 51–105.

[49] Kelton, W. Statistical analysis of simulation outputs. In *Proceedings of the 1997 Winter Simulation Conference* (1997), pp. 22–30.

[50] Kelton, W., and Barton, R. Experimental design for simulation. In *Proceedings of the 2003 Winter Simulation Conference* (2003), pp. 59–65.

[51] Kempf, K. Manufacturing planning and scheduling : where we are and where we need to be. In *Proceedings of the 5th Conference on Artificial Intelligent Applications* (1989), pp. 13–19. Invited presentation.

[52] Kim, K. Positioning of automated guided vehicles in a loop layout to minimize the mean vehicle response time. *International Journal of Production Economics 39* (1995), 201–214.

[53] Kim, Y., Kim, J., Choi, B., and Kim, H. Production scheduling in a semiconductor wafer fabrication facility producing multiple product types with distinct due dates. *IEEE Transactions on Robotics and Automation 7*, 5 (2001), 589–598.

[54] Knust, S. *Shop-scheduling problems with transportation.* PhD thesis, Universität Osnabrück, Germany, 1999.

[55] Koff, G. Automatic guided vehicle systems : Applications, controls and planning. *Material Flow 4* (1987), 3–16.

[56] Korte, B., and Vygen, J. *Combinatorial optimization : Theory and algorithms.* Springer, 2002.

[57] Kumar, P. Scheduling semiconductor manufacturing plants. *IEEE Control Systems Magazine 14*, 6 (1994), 33–40.

[58] Kumar, S., and Kumar, P. Queueing network models in the design and analysis of semiconductor wafer fabs. *IEEE Transactions on Robotics and Automation 17*, 5 (2001), 548–561.

[59] Kuo, C. Modelling and performance evaluation of an overhead hoist transport system in a 300mm fabrication plant. *International Journal of Advanced Manufacturing Technology 20* (2002), 153–161.

[60] Kusiak, A., and Haragu, S. The facility layout problem. *European Journal of Operational Research 29*, 3 (1987), 229–251.

[61] Lacomme, P. *Méthodes exactes et approchées pour l'optimisation des systèmes à moyens de transport.* Habilitation à Diriger des Recherches, Volume 1, Université Blaise Pascal Clermont-Ferrand II, France, 2005.

[62] Lacomme, P., Moukrim, A., and Tchernev, N. Simultaneously job input sequencing and vehicle dispatching in a single vehicle agvs : a heuristic branch and bound approach coupled with a discrete events simulation model. *International Journal of Production Research 43*, 9 (2005), 1911–1942.

[63] Lacomme, P., Tchernev, N., and Chu, C. An efficient framework for job input sequencing and vehicle dispatching in a flexible manufacturing system based on agv transport. In *Proceedings of the 7th IEEE International Conference on Emerging Technologies and Factory Automation* (1999), vol. 1, pp. 653–662.

[64] Law, A., and Kelton, W. *Simulation modeling and analysis.* Mc Graw Hill, New York, 2000.

[65] Le-Anh, T., and Koster, M. D. A review of design and control of automated guided vehicle systems. Tech. Rep. ERS-2004-030-LIS, Erasmus Research Institute of Management, Rotterdam School of Management / Rotterdam School of Economics, Erasmus Universiteit Rotterdam, The Netherlands, 2004.

[66] Leroudier, J. *La simulation à événements discrets.* AFCET, Hommes et Techniques, 1980.

[67] Liao, D., Jeng, M., and Zhou, M. Petri net modeling and lagrangian relaxation approach to vehicle scheduling in 300mm semiconductor manufacturing. In *Proceedings of the 2004 IEEE International Conference on Robotics and Automation*, pp. 5301–5306.

[68] Liao, D., and Wang, C. Neural-network-based delivery time estimates for prioritized 300-mm automatic material handling operations. *IEEE Transactions on Semiconductor Manufacturing 17*, 324–332.

[69] Lin, J., Wang, F., and Wu, C. The connecting transport of automated material handling system in wafer fab. *International Journal of Production Research 41*, 3 (2003), 529–544.

[70] Lin, J., Wang, F., and Yen, P. Simulation analysis of dispatching rules for an automated interbay material handling system in wafer fab. *International Journal of Production Research 39*, 6 (2001), 1221–1238.

[71] Lin, J., Wang, F., and Yen, P. Simulation analysis of the connecting transport amhs in a wafer fab. *IEEE Transactions on Semiconductor Manufacturing 16*, 3 (2003), 555–564.

[72] Lin, S., and Huang, H. Modeling and emulation of a furnace in ic fab based on colored-timed petri net. *IEEE Transactions on Semiconductor Manufacturing 11* (1998), 410–420.

[73] Little, J. A proof for the queuing formula $l = \lambda w$. *Operations Research 16* (1961), 651–665.

[74] Liu, H., Fung, R., and Jiang, Z. Modelling of semiconductor wafer fabrication systems by extended object-oriented petri nets. *International Journal of Production Research 43*, 3, 471–495.

[75] Mackulak, G., Lawrence, F., and Rayter, J. Simulation analysis of 300mm intrabay automated vehicle capacity alternatives. In *Proceedings of the 1998 IEEE/SEMI Advanced Semiconductor Manufacturing Conference* (1998), pp. 445–450.

[76] Mackulak, G., and Savory, P. A simulation-based experiment for comparing amhs performance in a semiconductor fabrication facility. *IEEE Transactions on Semiconductor Manufacturing 14*, 3 (2001), 273–280.

[77] Martin, D. Total operational efficiency (toe). In *Proceedings of the 1999 IEEE/SEMI Advanced Semiconductor Manufacturing Conference* (1999), pp. 37–41.

[78] Maxwell, W., and Muckstadt, J. Design of automated guided vehicle systems. *IIE Transactions 14*, 2 (1982), 114–124.

[79] Maza, S., and Castagna, P. Conflict-free agv routing in bi-directional network. In *Proceedings of the 2001 IEEE International Conference on Emerging Technologies and Factory Automation* (2001), pp. 761–764.

[80] Minsky, M. Models, minds and machines. In *Proceedings of the IFIP Congress* (1965), pp. 45–49.

[81] Montgomery, D. *Diseño y análisis de experimentos.* John Wiley & Sons, 1991.

[82] Montoya-Torres, J. Transport automatisé dans les systèmes de fabrication de semi-conducteurs. In *Actes des Journées Doctorales du GdR MACS* (2005). CD-ROM.

[83] Montoya-Torres, J. A literature survey on design approaches and operational issues of automated wafer-transport systems for wafer fabs. *Production Planning & Control* (forthcoming). Accepted in 2005.

[84] Montoya-Torres, J. Manufacturing performance evaluation in wafer semiconductor factories. *International Journal of Productivity and Performance Management* (forthcoming). Accepted in 2005.

[85] Montoya-Torres, J., and Campagne, J. Towards an optimised approach to control amhs in wafer manufacturing plants. In *Preprints of the Third IFAC/IEEE Conference on Management and Control of Production and Logistics* (2004), pp. 117–122.

[86] Montoya-Torres, J., Campagne, J., and Marian, H. Modélisation d'une unité de fabrication microélectronique. *Journal of Decision Systems* (submitted).

[87] Montoya-Torres, J., Dauzère-Pérès, S., Marian, H., and Vermariën, L. Analysis of transport strategies for full-automated semiconductor fabrication using large-scale facility models. In *Actes du 6e Congrès International de Génie Industriel* (2005). CD-ROM paper code 103-CIGI2005.

[88] Montoya-Torres, J., Dauzère-Pérès, S., and Vermariën, L. A simulation-optimization approach for vehicle control in automated semiconductor manufacturing. In *Proceedings of the 17th European Modeling Simulation Symposium* (2005), pp. 137–142.

[89] Montoya-Torres, J., Dauzère-Pérès, S., Vermariën, L., and Marian, H. Vehicle positioning in complex automated transport systems. In *Proceedings of the 10th IEEE International Conference on Emerging Technologies and Factory Automation* (2005), vol. 2, pp. 165–170.

[90] Montoya-Torres, J., Dauzère-Pérès, S., Vermariën, L., and Marian, H. Zone-based vehicle control in unified amhs for wafer fabs. In *12th IFAC Symposium on Information Control Problems in Manufacturing* (submitted). May 17-19, 2006. Saint-Etienne, France.

[91] Montoya-Torres, J., Vermarien, L., Campagne, J., and Marian, H. Design and operation of automated material handling systems for ic wafer semiconductor manufacturing. In *Proceedings of the 2nd International Industrial Simulation Conference* (2004), pp. 175–179.

[92] Montoya-Torres, J., Vermarien, L., Campagne, J., and Marian, H. Modèle conceptuel d'une unité de fabrication microélectronique. In *Actes de la 5e Conférence Francophone de Modélisation et Simulation "Modélisation et Simulation pour l'Analyse et l'Optimisation des Systèmes Industriels et Logistiques"* (2004), vol. 2, pp. 703–710.

[93] Moore, G. Cramming more components onto integrated circuits. *Electronics Magazine 38*, 8 (1965), 114–117.

[94] Moore, G. Lithography and the future of moore's law. In *Proceedings of the SPIE Conference on Optical/Laser Microlithography* (1995), pp. 2–17.

[95] Murata, T. Petri nets : properties, analysis and applications. *Proceedings of the IEEE*, 77 (1989), 541–580.

[96] Nance, R. Model representation in discrete event simulation : The conical methodology. Tech. rep., Department of Computer Science, Virginia Polytechnic Institute and State University, USA, 1981.

[97] Overwater, R. *Process and interactions, an approach to the modelling of industrial systems.* PhD thesis, Technische Universiteit Eindhoven, The Netherlands, 1987.

[98] Peters, B., and Yang, T. Integrated facility layout and material handling system design in semiconductor fabrication facilities. *IEEE Transactions on Semiconductor Manufacturing 10*, 3 (1997), 360–369.

[99] Pierce, N., and Stanfford, R. Modeling and simulation of material handling for semiconductor wafer fabrication. In *Proceedings of the 1994 Winter Simulation Conference* (1994), pp. 900–906.

[100] Pillai, D. Material handling automation for wafer fabrication facilities. In *Proceedings of the 1990 IEEE/CHMT International Electronic Manufacturing Technological Symposium* (1990), pp. 277–286.

[101] Qiu, L., Hsu, W., Huang, S., and Wang, H. Scheduling and routing for agvs : A survey. *International Journal of Production Research 40*, 3 (2002), 745–760.

[102] Rivero, C. *Contraintes mécaniques induites par les procédés de la microélectronique : développement des contraintes lors des réactions Co-Si et Ni-Si.* PhD thesis, Université Paul Cezanne Aix-Marseille III, France, 2005.

[103] Rodriguez-Verjan, G. Simulation analysis of amhs control policies in automated semiconductor manufacturing. Tech. rep., 300mm Manufacturing Programs, STMicroelectronics. Internship : INSA de Lyon, France, and Pontificia Universidad Javeriana, Bogota, Colombia, 2005.

[104] Rooda, J. *Modelling Industrial Systems.* Lecture Notes. Technische Universiteit Eindhoven, The Netherlands, 2000. http ://se.wtb.tue.nl/documentation/.

[105] Rooney, M., van der Burgt, M., Smit, T., and Rooda, J. Capacity calculation of an agv system in a mp2 wafer fab by means of simulation. poster. In *Proceedings of the 2001 International Symposium on Semiconductor Manufacturing* (2001).

[106] Saenz, C., Paternina-Arboleda, C., and Kathor, S. Simulation analysis of multi-load agvs. In *Proceedings of the Flexible Automation and Intelligent Manufacturing Conference* (1999), pp. 551–562.

[107] Sahni, S., and Gonzalez, T. P-complete approximation problems. *Journal of the ACM 23* (1976), 555–565.

[108] Schruben, L., and Yücesan, E. Transforming petri nets into event graph models. In *Proceedings of the 1994 Winter Simulation Conference* (1994), pp. 560–565.

[109] Smit, H. A hierarchical control architecture for job-shop manufacturing systems, 1992.

[110] Stalk, G., and Hout, T. *Competing Against Time : How Time-Based Competition is Reshaping Global Markets.* The Free Press, 1990.

[111] Sturm, R., Seidelmann, J., Dorner, J., and Reddig, K. An approach to robust layout planning of amhs. In *Proceedings of the 2003 Winter Simulation Conference* (2003), pp. 1366–1372.

[112] Tanchoco, J., and Egbelu, P. Potentials of bi-directinal guide-path for automated guided vehicles based systems. *International Journal of Production Research 24*, 5 (1986), 1075–1097.

[113] Ting, J., and Tanchoco, J. Unidirectional circular layout for overhead material handling systems. *International Journal of Production Research 38*, 16 (2000), 3913–3935.

[114] Ting, J., and Tanchoco, J. Optimal bidirectional spine layout for overhead material handling systems. *IEEE Transactions on Semiconductor Manufacturing 14*, 1 (2001), 57–64.

[115] Trebilcock, B. Overhead handling. *Modern Materials Handling 59*, 8 (2004), 43–47. Available on-line at http ://www.mmh.com.

[116] Tunali, S. Simulation for evaluating machine and agv scheduling rules in an fms environment. In *Proceedings of the 1995 IEEE Annual International Engineering Management Conference* (1995), pp. 433–438.

[117] Ulusoy, G., and Bilge, . Simultaneous scheduling of machines and automated guided vehicles. *International Journal of Production Research 31*, 12 (1993), 2857–2873.

[118] Uzsoy, R., Lee, C., and Martin-Vega, L. A review of production planning and scheduling models in the semiconductor industry. part i : System characteristics, performance evaluation and production planning. *IIE Transactions 24*, 4 (1992), 47–60.

[119] Uzsoy, R., Lee, C., and Martin-Vega, L. A review of production planning and scheduling models in the semiconductor industry. part ii : Shop-floor control. *IIE Transactions 26*, 5 (1994), 44–55.

[120] van Campen, E. *Design of a multi-process multi-product wafer fab.* PhD thesis, Technische Universiteit Eindhoven, The Netherlands, 2001.

[121] van der Meer, J. *Operational control of internal transport.* PhD thesis, Erasmus Universiteit Rotterdam, The Netherlands, 2000.

[122] Wang, M., Chung, H., and Wu, H. The evaluation of manual foup handling in 300-mm wafer fab. *IEEE Transactions on Semiconductor Manufacturing 16* (2003), 551–554.

[123] Wein, L. Scheduling semiconductor wafer fabrication. *IEEE Transaction on Semiconductor Manufacturing 1*, 3 (1988), 115–130.

[124] Wolf, S., and Tauber, R. *Silicon processing for the VLSI era : Volume 1 - process technology.* Lattice Press, 1986.

[125] Wood, S., and Saraswat, K. Modeling the performance of cluster-based fabs. In *Proceedings of the 1991 IEEE/SEMI International Semiconductor Manufacturing Science Symposium* (1991), pp. 8–14.

[126] Xiong, H., and Zhou, M. Scheduling of semiconductor test facilities via petri nets and hybrid heuristic search. *IEEE Transactions on Semiconductor Manufacturing 11*, 3 (1998), 384–393.

[127] Yang, T., and Peters, B. A spine layout design method for semiconductor fabrication facilities containing automated material handling systems. *International Journal of Operations & Production Management 17*, 5-6 (1997), 490–501, 426–427.

[128] Yang, T., Rajasekharan, M., and Peters, B. Semiconductor fabrication facility design using a hybrid search methodology. *Computers & Industrial Engineering 36* (1999), 656–583.

[129] Zant, P. V. *Microchip Fabrication : A Practical Guide to Semiconductor Processing.* McGraw-Hill, 2000.

[130] Zeigler, B. *Multifaceted Modeling and Discrete Event Simulation.* Academic Press, 1984.

Transport automatisé dans les systèmes de fabrication de semi-conducteurs : Nouvelles approches de gestion tactique et opérationnelle

Résumé : La fabrication de plaquettes de silicium (wafers) est l'un des processus de fabrication le plus complexes qui existe aujourd'hui. Étant donné à la fois le niveau très élevé d'investissements en équipement et la forte concurrence sur le marché, les fabricants de semi-conducteurs cherchent à rentabiliser au maximum les unités de fabrication (fabs) afin d'obtenir à la fois le meilleur niveau d'utilisation des équipements, un bas niveau d'encours et des temps de cycle de plus en plus courts. Avec l'arrivée de la nouvelle génération de wafers de 300mm de diamètre, des contraintes ergonomiques et de productivité imposent l'automatisation de toute la manutention et le transport des lots de production. Le système automatique de transport joue donc désormais un rôle essentiel pour l'amélioration de la productivité de la fab et la réduction des coûts de production. Cependant, les stratégies de pilotage du système de transport ont été relativement peu étudiées dans la littérature et les travaux existant le considèrent comme étant une entité isolée, ignorant ainsi les effets sur la gestion de la production et des équipements. Du fait de la complexité du processus de fabrication, les études proposées dans la littérature ont traditionnellement été réalisées avec de la simulation à événements discrets. L'objectif de cette thèse est de proposer une méthodologie d'analyse et d'optimisation de la gestion du système automatisé de transport des lots, afin d'améliorer les indicateurs clés de la production. Nous proposons d'abord une modélisation hybride de la fab et du processus de fabrication. Pour cela, nous utilisons une méthodologie basée sur l'identification des "process" et de leurs interactions, ainsi qu'une représentation en réseaux de Petri colorés temporisés. Cette conceptualisation nous permet de mieux intégrer les différents composants dans un modèle de simulation. Ensuite, plusieurs stratégies pour la gestion du système de transport des lots (i.e. stratégies de dispatching et de positionnement des chariots) sont analysées. Cette étude est réalisée en couplant des techniques classiques de la Recherche Opérationnelle (i.e. programmation mathématique, méthodes heuristiques) avec le modèle de simulation de la fab. Ainsi, la problématique de gestion du système de transport est abordée à partir du niveau tactique à l'aide d'une démarche d'optimisation qui permet de trouver les valeurs optimales des paramètres du modèle de simulation. Ce dernier permet d'étudier le comportement dynamique du système et facilite l'implémentation des règles pour la gestion opérationnelle. La mise en œuvre de la démarche hybride optimisation-simulation proposée est analysée dans la structure complexe d'une unité de fabrication de wafers de la nouvelle génération de 300 mm de diamètre.
Mots Clés : Système automatisé de transport et manutention (AMHS), pilotage de véhicules, fabrication de semi-conducteurs, optimisation, simulation.

Automated transport in semiconductor manufacturing systems : Novel approaches for tactical and operational management

Abstract : Semiconductor manufacturing is one of the most complex production processes. This is in fact a high-tech environment characterised by huge fixed costs for the required machinery, equipment and tooling. In order to achieve a satisfactory level of return on shareholders' equity, wafer fabrication facilities (fabs) are usually run to obtain short cycle times, low work-in-process inventory levels, and high equipment utilisation rates. In modern wafer fabs, an Automated Material Handling System (AMHS) is mandatory in order to satisfy ergonomic constraints and productivity objectives. Although the AMHS has a very important role in wafer fabs, it has been little studied in the literature from an operational point of view (i.e. vehicle positioning and dispatching strategies). Existing works consider the AMHS as an isolated system and do not study the impact of the variability generated in the rest of the fab. Because of the complexity of the wafer fabrication process, traditional methodologies have used discrete-event simulation approaches. The objective of this dissertation is to propose a formal methodology for the analysis and optimisation of the AMHS in order to improve the fab productivity metrics. We first propose a hybrid model of the fab using the Process-Interaction Approach and coloured timed Petri nets. These modelling formalisms aid us to characterise the physical production system, its control system, and the production flow process, in order to better identify the components of the simulation model. Then, various strategies for the operation of the AMHS are compared. This study is performed by coupling classical Operational Research-based techniques (i.e. integer programming, heuristics) with the simulation model of the fab. The problem of vehicle positioning is approached from a tactical level so as to identify the optimal parameters of the simulation model. This last allows us to study the behaviour of the transport system and facilitates the implementation of operation vehicle control policies. Final application of these strategies is done on a fab of the new 300mm wafer generation.
Keywords : Automated material handling, vehicle control, wafer manufacturing, simulation, optimisation.