

AMELIORATION DE LA REACTIVITE EN GESTION D'ATELIER PAR LA SIMULATION DISCRETE DANS UN CONTEXTE DE COOPERATION ENTRE DONNEUR D'ORDRES ET SOUS-TRAITANT

J.P.Kieffer* , J.Arékion** , K.Ayadi*** , D.Baudoin****

Résumé. - Cet article présente une étude préalable à la modélisation des mécanismes de coopération entre les systèmes d'ordonnancement d'un donneur d'ordres et de son principal sous-traitant, dans un contexte de gestion juste-à-temps d'un processus de production complexe. La simulation discrète y est utilisée comme support d'outils d'aide à la décision et comme moyen d'affiner et de clarifier les conditions de négociation entre les deux partenaires. Cette étude examine en sus les possibilités d'améliorer la coopération en adaptant certains concepts issus des systèmes 'multi-agents'. L'application des premiers résultats de l'étude est en cours dans un cas industriel. Ce n'est cependant qu'à l'issue de la mise en oeuvre complète qu'une évaluation précise des bénéfices de cette approche en termes de réactivité pourra être donnée.

Mots-clés : gestion de production, ordonnancement d'atelier, simulation discrète, multi-agents

1. Introduction

Dans le contexte actuel, les entreprises industrielles sont amenées à entretenir des relations étroites de partenariat avec leurs sous-traitants. Cela suit deux tendances essentielles : une baisse générale et à long terme des taux d'intégration chez les grands donneurs d'ordres et une généralisation des modes de gestion industrielle en "juste-à-temps".

* Professeur à l'université d'Aix-Marseille III (IUSPIM), chercheur au Département de Recherche en Informatique, Automatique et Mécatronique (DIAM), Centre Scientifique de Saint-Jérôme, Marseille.

** Maître de Conférences à l'IUT de Marseille à l'université d'Aix-Marseille III, chercheur au DIAM.

*** Enseignant à l'université d'Aix-Marseille III (IUSPIM), chercheur au DIAM.

**** Responsable de la gestion de production. Société Atmel-ES2, Rousset.

Il en résulte un besoin de coordination de plus en plus important entre donneurs d'ordres et sous-traitants, souvent organisés en réseaux d'entreprises, dans un souci de respect strict de délais.

Les éléments qui sont présentés ici s'inscrivent dans cette problématique et suggèrent une approche basée sur la simulation pour améliorer l'efficacité et la réactivité du pilotage à court terme de deux ateliers dans une relation de type client-fournisseur orientée dans une perspective de partenariat.

Après avoir présenté la problématique d'ensemble et justifié l'intérêt de l'utilisation de la simulation discrète et d'une approche de type "multi-agents", nous décrivons la démarche de modélisation et les premiers résultats obtenus, puis nous montrerons comment les modèles ont été utilisés pour l'aide à la décision en gestion de production. Enfin, nous aborderons l'extension de cette approche pour la mise en oeuvre d'un système de coopération entre les ordonnancements d'atelier du donneur d'ordres et de son sous-traitant.

2. Problématique d'ensemble

2.1 Spécificités et difficultés de la gestion de production dans l'industrie du semi-conducteur

Notre réflexion concerne le secteur des industries des semi-conducteurs et s'appuie sur le cas de l'entreprise Atmel-ES2, qui produit sur commande et sur programme des circuits intégrés spécifiques (ASIC) dont le processus de production est schématisé par la figure 1.

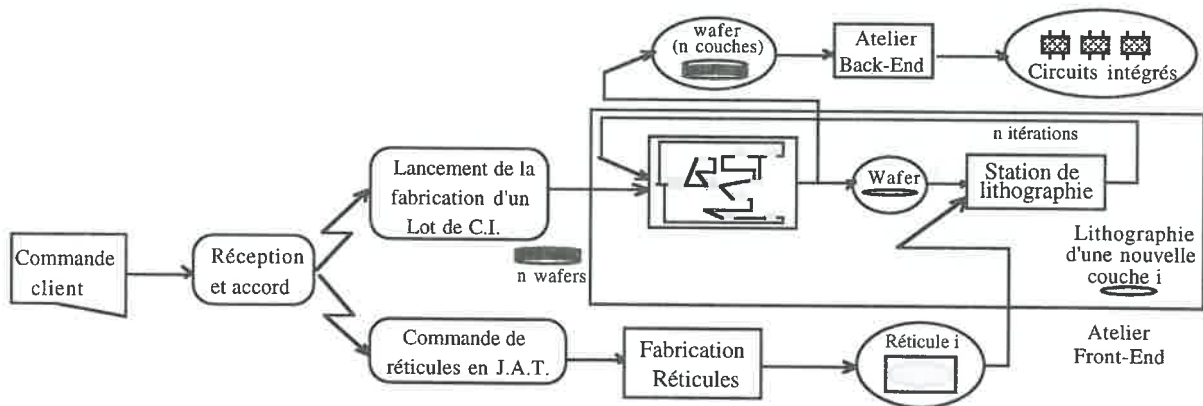


Figure 1 : Processus de production des circuits intégrés

Ce processus de production est l'un des plus complexes que l'on peut trouver dans la classe des systèmes de production de type "job shop". Il est composé de deux étapes principales. Dans un atelier amont ("front-end"), les multiples couches composant les circuits intégrés sont réalisées sur une plaque de silicium dénommée "wafer".

Dans un atelier aval ("back-end"), les wafers sont ensuite découpés pour obtenir les circuits intégrés proprement dits et ces derniers sont assemblés avec leurs éléments de protection et de connectique, puis contrôlés et conditionnés. L'atelier "back-end", qui ne pose

pas de problèmes particuliers de gestion de production et qui ne nécessite pas de coopération entre donneur d'ordres et sous-traitant, ne sera pas examiné dans la suite de cet article.

Dans la gamme de fabrication des wafers, qui peut compter jusqu'à 400 opérations élémentaires, apparaissent des opérations critiques sur des machines coûteuses, dont notamment des machines de lithographie optique nécessitant des réticules qui sont sous-traités.

Suite à une commande client, le service de gestion de production du donneur d'ordres lance simultanément le début de la fabrication d'un lot de wafers et une commande de réticules à son sous-traitant. Ces réticules doivent être disponibles au moment prévu de passage du lot en lithographie. D'autres opérations de fabrication succèdent à la lithographie. Ce processus est réitéré autant de fois que le circuit intégré comporte de couches (nombre de couches moyen : 12). A chaque couche correspond un réticule spécifique. Les dates de livraison des réticules sont donc échelonnées dans le temps en fonction des dates prévues pour les passages successifs du lot de wafers en lithographie. La réalisation de tout ce processus nécessite un travail effectif de 10 jours environ, alors que le temps de cycle est de trois semaines.

Ce processus nécessite donc :

- une estimation précise, pour chaque lot, du temps de cycle nécessaire pour décrire la boucle de réitération, afin de prévoir le plus finement possible les dates de passage des lots par la phase de lithographie;
- un bon respect par le sous-traitant de la série de dates de livraison des réticules en lithographie.

En effet, compte tenu des nombreux bouclages dans la fabrication, un défaut de synchronisation entre l'arrivée d'un réticule et du lot correspondant en lithographie entraîne une forte propagation de perturbations dans l'ensemble du système de production. Si ce dysfonctionnement se produit, une remise en cause du plan de production du donneur d'ordres ou du sous-traitant, avec éventuellement une concertation rapide entre eux, est nécessaire pour rétablir la situation par des changements de priorités d'ordonnement de part et d'autre.

De plus, des modifications fréquentes de quantités peuvent intervenir de la part des clients. Une planification à long terme étant difficile à réaliser en raison de l'horizon court du carnet de commandes et de la variabilité de ces dernières, ces perturbations commerciales sont à absorber au niveau de l'ordonnement d'atelier des deux partenaires, puisqu'elles concernent principalement l'étape de lithographie. Cette étape nécessite entre 0.4 et 1.8 heures de travail, alors que le temps moyen d'attente devant cette station est d'environ 3 heures. L'un des principaux intérêts de cette étude est donc de réduire ces pertes de temps.

Le problème qui vient d'être décrit, à savoir la nécessité de synchronisation entre un flux principal de produits et un flux secondaire provenant d'un sous-traitant est très fréquent dans le monde industriel. Cependant, il prend ici une dimension particulière du fait des nombreux bouclages de flux qui viennent cumuler les contraintes de synchronisation et additionner les décalages temporels sur ces flux. Au delà des efforts d'optimisation des

performances de production internes à chaque entreprise, il est donc nécessaire de mettre en oeuvre une coopération dynamique pour favoriser une harmonisation des plans de production du donneur d'ordre et de son sous-traitant.

2.2 *Intérêt de la simulation discrète pour l'aide à l'ordonnancement*

L'intérêt d'utiliser ici la simulation discrète tient à l'impossibilité de mettre en oeuvre une technique fondée sur un diagramme de GANTT : nombre important de tâches à placer (gammas de fabrication comportant plus d'une centaine d'opérations), multiples réitérations de certaines opérations et durées de tâches pouvant être aléatoires pour des raisons de disponibilité d'équipements soumis à des défaillances. Cela a aussi pour conséquences de rendre prohibitif le temps de calcul d'un algorithme d'ordonnancement, d'interdire une visualisation ergonomique du résultat de placement, et de rendre peu fiables les dates prévisionnelles calculées.

Compte tenu des limites de l'ordonnancement classique évoquées au paragraphe précédent, nous nous sommes orientés vers l'utilisation des techniques de simulation discrète, comme support d'outils d'aide à la décision. En effet :

- tout d'abord, cette approche, basée sur la description du système de production en termes de serveurs et de files d'attente, est conforme à la pratique courante dans l'industrie du semi-conducteur, où les lots de wafers s'accumulent devant les machines et sont opérés en fonction de règles de priorité ;
- ensuite, la simulation permet de s'affranchir des limites de l'ordonnancement par affectation formelle de tâches à des ressources ; elle permet en effet de tester des règles de décision "opportunistes" au niveau des priorités d'ordonnancement ;
- enfin, compte tenu de la complexité des problèmes de pilotage traités ici, il semble judicieux de responsabiliser les opérateurs en leur réservant une place active dans les prises de décision, plutôt que d'automatiser fortement ces dernières. L'utilisation de la simulation discrète comme aide à la décision entre bien dans ce schéma, puisqu'elle permet de définir des cadres de décision indicatifs composés de règles de priorité suggérées dont l'efficacité aura été testée sur un modèle.

Par ailleurs, on notera que ce choix de la simulation discrète dans ce contexte est cohérent avec des approches déjà proposées, notamment par (Dayoff et Atherton, 1984), (Wien, 1988), ainsi que (Glasse et Resende, 1988).

2.3 *Intérêt d'une approche de type "multi-agents" pour améliorer la coopération entre donneur d'ordres et sous-traitant*

L'approche multi-agents, telle qu'elle est conçue en intelligence artificielle distribuée (Ferber J., 1989) semble, pouvoir être transposée et appliquée avec intérêt dans le cas d'entités d'ordonnancement coopératives, comme cela a été montré dans (Baillet P, 1994).

Dans cette dernière publication, notre équipe de recherche a en effet envisagé une décentralisation des fonctions de décision basée sur une coordination entre des centres de décision d'ordonnancement, susceptible d'augmenter la réactivité du pilotage face à des perturbations.

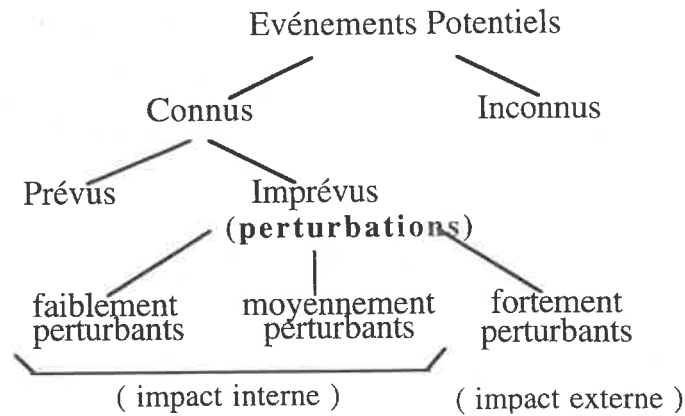


Figure 2 : Classification des perturbations

Nous y avons également proposé une classification des perturbations (Cf. figure n° 2) et détaillé les concepts de hiérarchisation des traitements de ces perturbations, ainsi que les procédures de communication entre centres (Cf. figure n° 3). Sur cette dernière figure, on remarquera la correspondance entre les aiguillages et les niveaux de perturbation (D 11 et D12 pour les perturbations faibles, D 13 pour les perturbations moyennes, et D 2 pour les perturbations fortes).

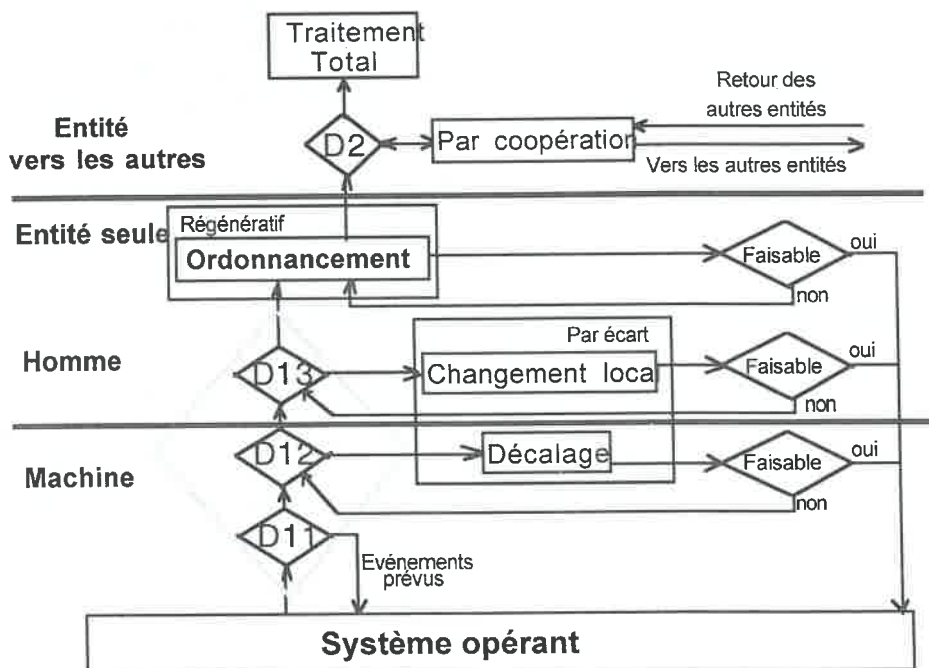


Figure 3 : Organisation des traitements progressifs des perturbations

Dans le cas présent, les mécanismes de coopération présentés dans (Baillet P., 1994) peuvent être adaptés selon le principe décrit figure n° 4. Ainsi, dans le cas d'une perturbation détectée au niveau du système opérant du donneur d'ordres, une suite d'aiguillages (D11 à D14) permet de déterminer un niveau de réaction adapté à l'importance de la perturbation. Suivant le cas, la réaction peut être directe, peut passer par un dialogue avec le sous-traitant,

ou encore peut nécessiter des remises en cause plus globales du plan de production ou des délais clients. Les décisions prises au niveau des aiguillages supérieurs sont éclairées par un système d'aide à la décision basé sur la simulation et pouvant être utilisé dans deux circonstances correspondant respectivement à l'activation des modules "simulation" 11 et 12 de la figure 4. Les premiers niveaux de réaction du sous-traitant (D 21 et D 22) mettent en jeu un mécanisme similaire.

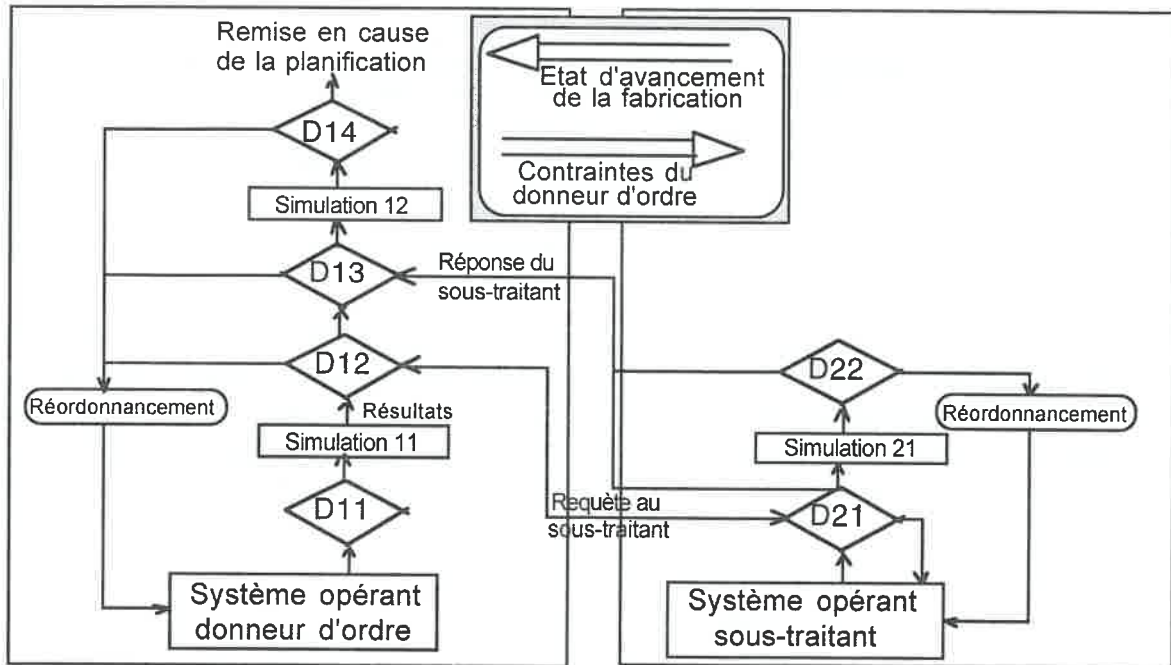


Figure 4 : Mécanisme de coopération entre donneur d'ordres et sous-traitant

3. Conception et mise en oeuvre d'un simulateur

A ce stade du travail, a été construit en laboratoire un modèle global fonctionnant "hors ligne", représentatif de l'atelier "front-end" du donneur d'ordres et destiné à valider la démarche. Ce modèle a par ailleurs été destiné à être transposé par la suite sur le site industriel, pour l'intégrer dans un ensemble d'outils d'aide à la décision fonctionnant "en ligne", c'est-à-dire de manière connectée au système de suivi de production.

Dans ce paragraphe, nous ne décrivons de manière détaillée que les aspects concernant le premier modèle réalisé en laboratoire.

3.1 Méthodologie de conception et d'implantation du modèle de simulation

3.1.1 Objectifs de modélisation

Les principaux objectifs de modélisation ont été les suivants :

- représentation du système existant, prenant en compte les multiples flux différents de produits dans l'atelier, en faisant l'hypothèse, dans ce modèle, que les réticules sont toujours disponibles ;

- simplification du modèle par rapport au système modélisé, réalisée par le biais de l'agrégation d'un certain nombre de ressources non critiques ; et cela en vue d'éviter une complexité trop importante du modèle et de faciliter la compréhension et l'interprétation des résultats de simulation ;
- modélisation des flux critiques pour évaluer l'application de règles de priorité opportunistes aux ressources critiques ;
- prise en compte de la disponibilité des équipements (dépendant des défaillances aléatoires, ainsi que des politiques de maintenance préventive ou curative), afin d'évaluer son incidence sur les temps de cycle et la productivité de l'atelier considéré dans son ensemble ;
- évaluation des conséquences de la politique de lancement des ordres de fabrication de wafers dans l'atelier en termes de temps d'attente devant les ressources critiques.

3.1.2 Méthodologie de modélisation

Le modèle a été supporté par l'outil de simulation SLAM³ et il a été réalisé selon la méthodologie d'affinage progressif, que nous avons déjà présentée dans (Dufrene, Gharbi et al. 1994) et (Dufrene, Kieffer, Rebouché, 1994). Ainsi, nous avons commencé par construire un modèle global du système, en posant un certain nombre d'hypothèses d'agrégation des ressources et des produits. Ce modèle a été affiné progressivement à des niveaux de détail de plus en plus fins. Les principaux avantages de cette méthode sont les suivants :

- cela permet une validation précoce de la stratégie de modélisation et des alternatives pertinentes de conduite de l'étude ;
- cela est cohérent avec la méthode SADT, qui convient bien pour l'analyse préalable du fonctionnement du système modélisé et pour la spécification du modèle à construire, et qui est organisée elle aussi en allant du global au détail ;
- la validation du modèle peut être progressive et des résultats immédiats peuvent être obtenus. Ces résultats intermédiaires permettent, le cas échéant, une réorientation de la modélisation sans à avoir à attendre la fin de l'étude.

La caractéristique la plus importante du modèle que nous avons construit est la représentation de flux de produits comportant de nombreuses variantes et de multiples itérations sur les mêmes ressources, comme le montre la figure n° 5.

Cette figure correspond à un modèle générique dans lequel des attributs sont assignés aux entités en fonction des différents types de flux et de process. Des tests de branchement sont effectués sur ces attributs pour représenter les différentes variantes de flux, telles que par exemple la réintroduction des entités en un point particulier du modèle, en fonction du type d'itération à considérer dans le processus de production.

Ainsi, par exemple, on peut remarquer sur la figure n° 5 que les produits en sortie de la station "THINOX" peuvent être orientés vers les stations "COAT" (n°1), "IMPLANT" (n° 3) ou POLYDEP (n°5), en fonction de leur type et de leur degré d'avancement dans le process.

³ Logiciel distribué par PRITSKER and ass., P.O. Box 2161, West Lafayette, Indiana 47906, USA.

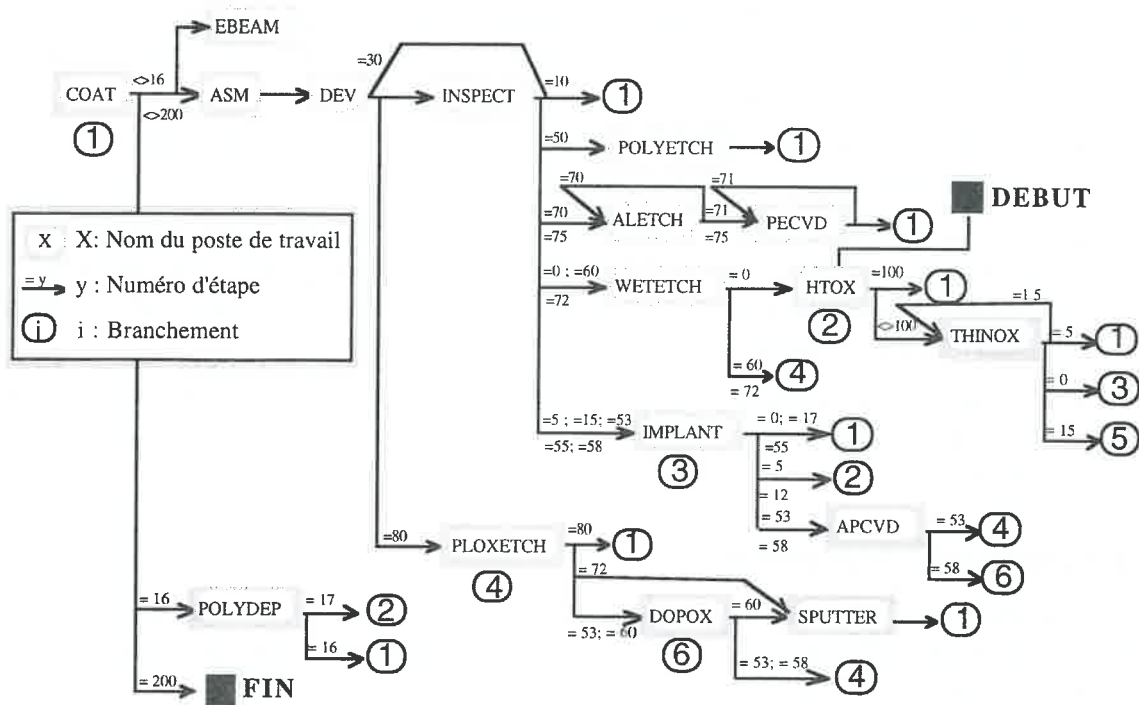


Figure 5 : Modèle du processus de production des wafers

Cela rend le modèle "flexible", puisque n'importe quel nouveau produit peut être aisément représenté et intégré, au moyen d'une simple modification des attributs, sans que cela oblige à modifier la structure du modèle.

3.1.3 Quelques aspects de l'implémentation du modèle

Le modèle a été construit à la suite d'une étude exhaustive des données de suivi de production collectées sur une durée de 3 mois (à comparer aux résultats de simulation obtenus sur une durée équivalente). C'est cette analyse qui a permis de poser les hypothèses simplificatrices suivantes, correspondant à la première phase de la méthode d'affinage progressif :

- nous n'avons retenu que deux principaux types de caractéristiques au niveau du processus de lithographie (ASM et EBEAM);
- nous avons agrégé les postes de travail en fonction de leur charge de travail observée (environ 30% d'entre eux ne sont pas critiques et peuvent être considérés comme étroitement liés aux autres postes et regroupés avec ceux-ci).

Il en résulte un modèle agrégé comportant 17 postes de travail (machines ou groupes de machines). Ces postes peuvent opérer un seul ou plusieurs lots simultanément.

Les défaillances des ressources sont représentées par une génération aléatoire d'entités virtuelles dans le modèle, entraînant des interventions de maintenance modélisées sous forme d'activités opérant ces entités virtuelles. A ce stade, nous avons considéré une distribution

uniforme de la génération aléatoire, cohérente avec le taux moyen de défaillance observé en pratique.

Chaque poste est modélisé par un sous-modèle standardisé. Les liens entre sous-modèles sont représentatifs des multiples itérations du process décrites figure n° 1. Ces liens sont gérés par les attributs mentionnés au paragraphe 3.1.2.

La génération d'entités représentant le lancement des ordres de fabrication est en conformité avec la pratique observée dans l'entreprise, à savoir :

- deux tailles de lots (25 % de petits lots et 75 % de grands) ;
- deux types de process (30 % EBEAM et 70 % optique) ;
- lancement des ordres de fabrication pour maintenir constant le niveau des en-cours dans l'atelier.

3.2 *Expérimentation du simulateur*

3.2.1 Validation du modèle

Cette phase s'est déroulée en deux étapes.

La première a consisté en une comparaison globale entre cycles de production réel et simulé, et cela pour chaque type de process. Pour ce faire, nous avons créé un ordre de fabrication unique d'un wafer, et observé sa trace dans la simulation. Comme on élimine de ce fait la formation de files d'attente d'entités devant les ressources, on peut comparer ce temps de cycle total simulé à la somme des temps opératoires de la gamme, qui est d'environ 180 heures.

Dans la seconde étape, nous avons simulé le flux de production complet, en prenant en compte toutes les contraintes existantes. Le tableau n° 1 montre, pour chaque station, les ratios charge/capacité observés et simulés. Les écarts relativement faibles entre ces ratios traduisent une assez bonne fiabilité du modèle, en dépit des approximations qui ont été faites à ce stade de l'étude.

On peut cependant remarquer sur ce tableau que le taux de chargement des postes les plus chargés (charge supérieure à 80 % de la capacité) est systématiquement sous-estimé par la simulation, notamment pour les postes ASM (lithographie nécessitant des réticules) et INSPECT. Cela s'explique surtout par l'hypothèse de disponibilité des réticules que nous avons posée (ces deux postes, pratiquement adjacents dans le processus, sont en effet les plus directement concernés par cette hypothèse).

Poste de travail	Réel	Simulé
HTOX	61%	60%
COAT	41%	40%
POLYDEP	60%	55%
ASM	91%	80%
EBEAM	75%	80%
DEV	37%	35%
INSPECT	97%	85%
POLETCH	7%	10%
ALETCH	47%	55%
PECVD	48%	55%
WETETCH	5%	10%
THINOX	81%	80%
PLOXETC	39%	55%
DOPOX	68%	60%
SPUTTER	51%	70%
IMPLANT	92%	90%
APCVD	51%	30%

Tableau n° 1 : Ratios charge/capacité pour chaque poste de travail:

3.2.2 Principaux résultats de la simulation

Les résultats ont été obtenus à la suite de multiples jeux d'essai correspondant à différentes hypothèses sur les contraintes, sur les paramètres des distributions statistiques des défaillances et des temps de réparation des ressources, ainsi que sur les règles de priorité utilisées (au niveau global pour le lancement des ordres de fabrication et au niveau local pour chaque poste de travail).

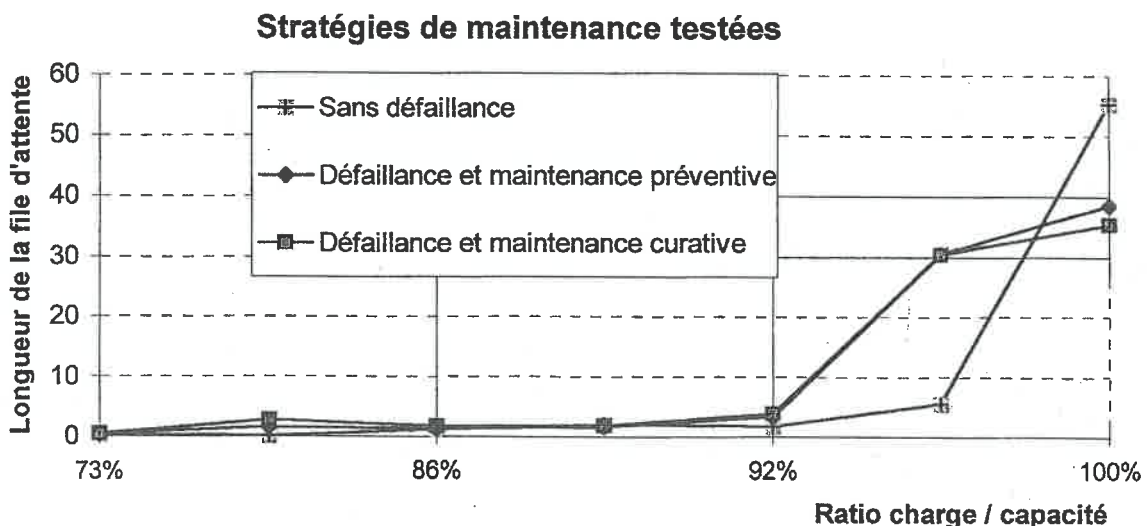


Figure 6 : Impact des politiques de maintenance

D'un point de vue global, on peut observer une relation directe entre, d'une part, la longueur moyenne des files d'attente dans l'atelier, et, d'autre part, le taux de défaillance ainsi que la politique de maintenance de chaque ressource. La figure n° 6 montre les résultats de trois hypothèses testées : l'hypothèse théorique d'une absence de défaillance sur les

ressources, ce qui est un cas idéal ; et, en cas de défaillance, l'application de stratégies de maintenance préventive, d'une part et, d'autre part, de maintenance curative. Cette étude illustre bien l'engorgement de l'atelier lorsque le ratio charge/capacité dépasse 90 % (et cela quelles que soient les hypothèses relatives aux défaillances et à la maintenance).

D'un point de vue plus détaillé, l'influence de quatre règles de priorité, sur les files d'attente devant les machines, a été analysée : FIFO (premier entré, premier sorti), priorité aux plus petits ou aux plus gros lots, et priorité aux lots urgents (priorité commerciale). Des comportements similaires à ceux décrits à l'alinéa précédent ont été observés. Ils sont illustrés par la figure n° 7. A première vue, ces résultats peuvent paraître assez évidents. Cependant, ils nous ont permis de déceler des transferts des files d'attente d'un goulot d'étranglement à un autre en fonction des politiques testées.

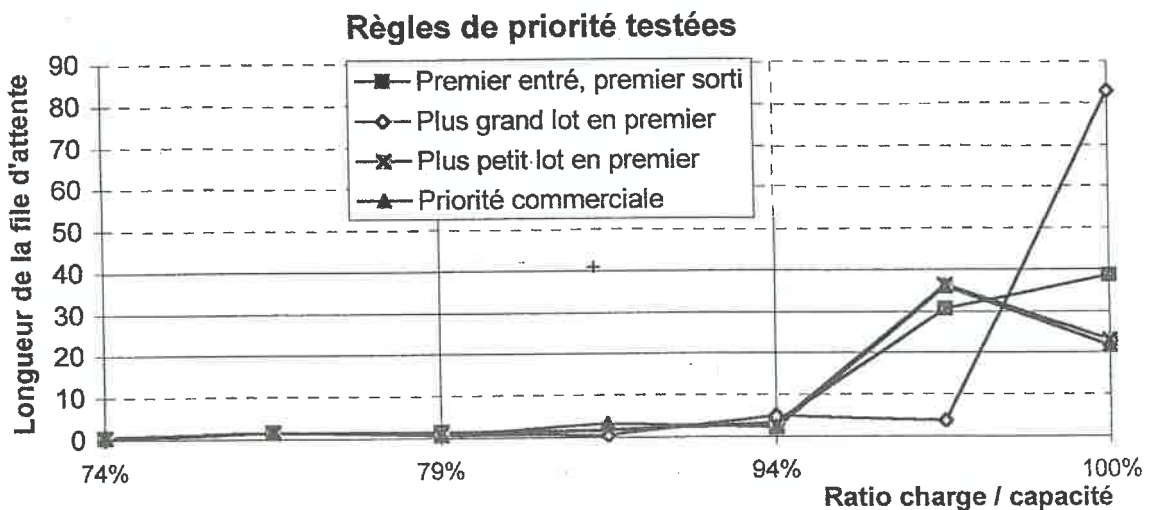


Figure 7 : Impact des règles de priorité

Nous pensons que ce phénomène résulte de la configuration du système de production ; dans la plupart des cas, les flux de produits sont structurés selon une séquence linéaire et on peut vérifier les hypothèses bien connues de la méthode OPT, c'est-à-dire que le débit d'un goulot d'étranglement limite le débit global de l'atelier. Ici, a contrario, la complexité des flux de produits, avec de multiples itérations et de multiples variantes, crée un réseau complexe entre les postes de travail. Il est par conséquent inutile d'optimiser le débit des postes les plus chargés, dès lors que les produits iront s'accumuler devant n'importe quel poste critique du process.

4. Exploitation du simulateur pour l'aide à la décision en gestion de production

Le premier modèle du système de production du donneur d'ordres a été fiabilisé et permet de prévoir avec une bonne précision les comportements de ce système. Ses principales caractéristiques et les résultats obtenus ont été présentés dans (Ayadi K. et al., 1995) et résumés au paragraphe précédent.

Sur la base de cette première étude de faisabilité, un modèle similaire a été bâti en utilisant le progiciel MANSIM. Ce modèle de simulation de l'atelier "front-end" existant est directement alimenté par les données de suivi de production du donneur d'ordres et permet ainsi une aide à la décision "en ligne". Il permet notamment de prévoir avec une meilleure finesse les dates de passage des lots en lithographie, et donc les dates de livraison de réticules à demander au sous-traitant.

Cette meilleure finesse donne maintenant des écarts entre données observées et données simulées inférieurs à 5 % quel que soit le poste de travail considéré, ces écarts ne conduisant plus comme précédemment à une minoration systématique du taux de charge simulé des postes critiques. Une telle amélioration par rapport aux résultats du modèle "hors-ligne" a été obtenue en prenant des hypothèses plus proches de la réalité observée, et notamment :

- en modélisant de manière plus précise les défaillances des ressources (taux de disponibilités personnalisés par ressource au lieu d'un taux de disponibilité moyen) ;
- en introduisant dans le modèle des taux de rebuts correspondant à l'exploitation statistique des résultats observés du contrôle de qualité ;
- en affinant la représentation des règles de décision.

Ce second modèle a en effet été élaboré sur la base du premier pour représenter aussi fidèlement que possible les processus de décision dans l'atelier. Pour ce faire, les règles de décision et de priorité ont été affinées et complétées en prenant en compte les aspects décrits ci-après.

Les processus de décision afférents au pilotage de l'atelier peuvent être classés en deux catégories, correspondant au découpage des responsabilités dans l'entreprise : d'une part, les processus de décision globaux concernant l'ensemble des postes de travail et l'atelier pris dans son ensemble, et, d'autre part, les processus de décision locaux, qui ne mettent en cause qu'un poste ou un groupe de postes de travail particulier, considérés souvent comme critiques (goulots d'étranglement).

Les décisions globales résultent de simples algorithmes destinés à obtenir des résultats d'ensemble sur certains objectifs tels que le respect des délais clients. Les décisions locales, au contraire, mettent en jeu des heuristiques faisant intervenir des règles d'ordonnancement opportunistes en fonction des situations.

4.1 Formalisation des règles de décision globales

A ce niveau, le responsable a à se préoccuper des questions suivantes:

- minimisation du temps de cycle,
- réduction des retards de livraison,
- équilibrage de la charge entre les différentes zones de l'atelier,
- optimisation de la charge sur les goulots d'étranglement,
- maximisation à court terme du flux financier facturable.

Pour atteindre ces objectifs, il est nécessaire de lancer les ordres de fabrication en appliquant les règles suivantes :

- a) les ordres de fabrication sont divisés en deux catégories, suivant que leur gamme les amène à être opérés sur l'un ou l'autre des deux goulots d'étranglement suivants: lithographie optique ou par procédé EBEAM ;
- b) les ordres de fabrication ayant une valeur ajoutée importante sont lancés prioritairement ;
- c) les ordres de fabrication sont lancés par lots, et on en lance suffisamment chaque jour pour être certain de ne pas créer de ruptures de charge sur les ressources critiques, suivant de ce point de vue les recommandations de la méthode OPT ;

La politique de lancement est établie en fonction des charges sur les postes de lithographie (Lawton et al. 1990). Un niveau de charge constant (Steve et al. 1994, Glassey 1988) est appliqué quotidiennement à chaque goulot d'étranglement.

Cette politique de lancement n'est pas appliquée pour des ordres de fabrication de circuits spécifiques qui sont lancés avec une priorité maximale. Et cela de manière à satisfaire la règle b) ci-dessus, puisque ces ordres correspondent à des fabrications par petits lots à forte valeur ajoutée.

4.2 Formalisation des règles de décision locales

Dans le cadre général des décisions prises par le responsable de production, les responsables des baies (regroupement de postes de travail en îlots de fabrication) ont une certaine latitude pour gérer la charge des postes en appliquant deux types de règles :

- saturation de la capacité de certaines machines, par exemple en regroupant plusieurs lots de wafers quand ces lots sont de petite taille ;
- déclassement de la priorité d'un lot si celui-ci doit être ensuite envoyé dans une file d'attente devant une ressource saturée.

Cette latitude est cependant maintenue dans certaines limites car son incidence sur la stabilité des lignes de production n'est pas négligeable. En effet, on pourra assister à des phénomènes de "pompage" où des en-cours importants s'accumulent devant certains postes alors que d'autres postes sont inactifs.

4.3 Mise en oeuvre des règles de décision

Au **niveau global**, les règles de décision décrites au paragraphe 4.1. ont été implémentées dans le simulateur. Cependant, pour faciliter leur choix par le responsable en fonction des circonstances, des aides ont été élaborées.

Une série d'indicateurs a donc été développée en utilisant des méthodes graphiques, telles qu'elles sont définies dans (Lozinski et al. 1988 et Fargher et al. 1994). La figure n° 8 donne un exemple de ce type de graphique établi pour un poste critique particulier. Ainsi, sur un horizon de 8 jours, on trace, jour après jour pour diverses catégories de lots, la variation cumulée de l'en-cours, ce qui permet de comparer cette variation à la variation-objectif

représentée par la droite. Sur ce graphique, on remarque par exemple que, dans le cas illustré, le responsable a adapté progressivement sa politique de lancement de manière à obtenir une quantité d'environ 150 wafers à opérer sur le poste en question à la fin de l'horizon de 8 jours.

Par ailleurs, le calcul d'un "ratio critique"⁴ (Panwalkar et al. 1977, Wien et al. 1988) permet de répartir les lots en 9 classes de priorité et de les traiter en fonction de leur avance ou de leur retard par rapport au délai client prévu.

Enfin, un utilitaire de gestion des gammes de fabrication permet de moduler la politique générale de lancement pour certains postes de travail, notamment quand la capacité ou le temps de préparation du poste nécessite des tailles "économiques" de lots différentes des tailles habituellement adoptées sur la plupart des postes. Ainsi, pour les machines pouvant traiter un nombre important de wafers simultanément, l'utilitaire adapte la taille de la charge à traiter.

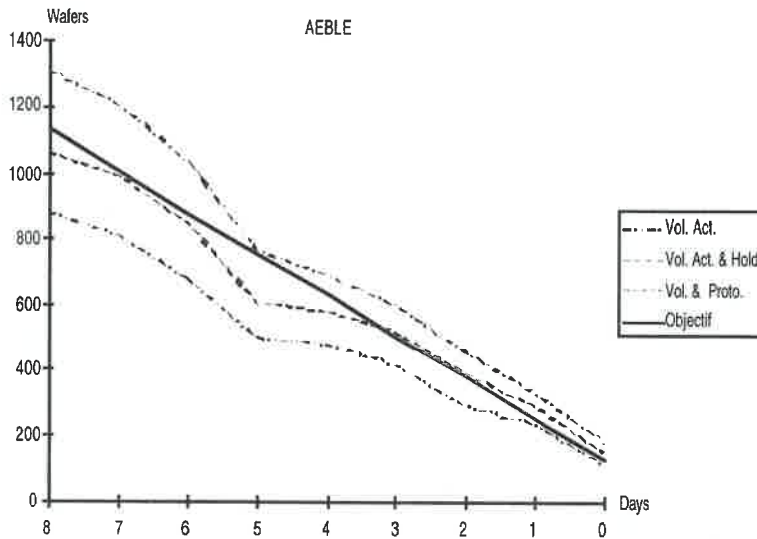


Figure 8 : Courbe décumulée de l'évolution des en-cours devant un poste de travail

Au **niveau local**, les responsables appliquent deux stratégies pour harmoniser la productivité des baies.

La ligne de production est divisée en 13 centres de charge. En établissant des règles de fonctionnement combinant le calcul du ratio critique et un certain nombre d'heuristiques, on cherche :

- d'une part, à réguler les temps de cycle avec une marge de 10% et à homogénéiser les niveaux des en-cours dans chaque centre de charge ;

⁴ La notion de ratio critique est classique en ordonnancement. De façon générale, elle exprime le rapport entre le temps nécessaire pour terminer un travail en cours (ou un lot en cours de fabrication) et le temps restant disponible pour cela. Elle peut être par exemple exprimée par le ratio (marge restante) / (nombre d'opérations qui restent à effectuer).

- d'autre part, à maintenir un niveau d'activité journalière identique dans chaque centre de charge de manière à obtenir des quantités équivalentes de wafers en sortie des différents centres, ce qui a pour but de réguler les flux sur l'ensemble de l'atelier.

Le simulateur est utilisé pour calculer un niveau d'activité standard à partir des données de production. Ce calcul est effectué chaque semaine en tenant compte de l'activité de la semaine précédente et des OF à traiter dans la semaine en cours. Ainsi, l'activité de chaque centre est comparée à ce standard et les priorités sont affectées aux centres les moins actifs.

De même, le niveau d'en-cours idéal devant chaque poste de travail est calculé à partir des données de suivi de production. Ce calcul part du principe que l'en-cours total de l'atelier doit être réparti dans les files d'attente devant les différents postes "équitablement", c'est-à-dire de manière inversement proportionnelle au temps de traitement des lots en attente devant eux. Cela permet ainsi de détecter les accumulations anormales de wafers devant un poste donné.

5. Vers la mise en oeuvre d'une approche de type "multi-agents"

L'ensemble des processus de décision qui vient d'être décrit est fortement affecté par la qualité des relations entre donneur d'ordres et sous-traitant, et mérite donc d'être complété par une amélioration de ces relations, dans l'esprit des suggestions qui ont été avancées au paragraphe 2.3.

Cette amélioration concerne à la fois la concertation et la coordination de planification et la réactivité à court terme aux perturbations susceptibles d'affecter la fabrication chez le donneur d'ordres et le sous-traitant. Pour la planification, la mise en place d'un échange de données informatisé permet d'améliorer l'estimation des délais clients prévisionnels, ainsi qu'une diminution des délais administratifs de transmission des ordres de l'un vers l'autre. L'utilisation hors ligne des simulateurs décrits précédemment permet également de prévoir les temps de cycle de manière plus fiable. Le second aspect, concernant la coopération pour une réaction aux perturbations, est celui qui nous intéresse ici.

Dans l'état actuel de la réalisation, les procédures de réordonnancement sont effectuées par l'homme, ainsi que les dialogues de coordination entre les responsables d'ordonnancement du donneur d'ordres et du sous-traitant. Ces dialogues sont en cours d'automatisation par l'installation d'un système d'échanges de données informatisés (EDI) entre les deux entreprises, de manière à réduire sensiblement les temps de transfert des informations et à augmenter la fréquence des coordinations.

Cet EDI transférera également du sous-traitant vers le donneur d'ordres les données de suivi de production des réticules en cours. Le modèle actuel supporté par MANSIM sera complété par un sous-modèle représentatif du processus de production des réticules. Ainsi, ces données alimenteront le simulateur qui fournira une information interprétée en termes d'avance ou de retard par rapport aux dates prévues de disponibilité des réticules sur le poste de lithographie. Cela permettra de lever l'hypothèse forte de disponibilité des réticules

qui avait jusqu'ici été à la base des modèles testés. Le donneur d'ordres pourra alors utiliser ce modèle complété pour prendre par avance des décisions de réordonnement.

Il faut en outre souligner l'intérêt d'exploiter les deux suivis par un seul modèle. Ainsi, lorsqu'un dysfonctionnement surviendra chez le donneur d'ordres, ce dernier pourra réagir en demandant au sous-traitant de modifier les priorités initiales qui lui avaient été données.

Dans une dernière étape, la démarche pourra être complétée par l'installation d'un modèle de simulation supplémentaire (Cf figure n° 4, module "Simulation 21"), ce qui pourra permettre au sous-traitant d'affiner ses réponses aux requêtes du donneur d'ordres.

Enfin, parallèlement à cette démarche de mise en oeuvre, les modèles de simulation seront exploités "hors ligne" pour définir et tester des règles de décision "opportunistes" au niveau des priorités d'ordonnement.

6. Conclusion et perspectives

Nous avons présenté ici une approche particulière d'utilisation de la simulation discrète pour l'aide à l'exploitation des systèmes de production. Cette approche se justifie ici par la complexité de la problématique d'ordonnement, qui rend délicate la mise en oeuvre de techniques classiques de placement de tâches. Cependant, dans d'autres cas moins complexes, la démarche de coopération entre centres de décision d'ordonnement pourrait tout aussi bien être supportée par une simulation reposant sur un algorithme de placement de tâches sur un diagramme de GANTT, ce qui ne changerait rien au principe exposé.

L'ensemble des problèmes et des solutions présentés dans cet article introduit une poursuite de recherches en simulation, vers la mise en oeuvre de modélisations de systèmes complexes sous forme de réseaux de modèles coopérants. Cela s'inscrit dans la logique de transformation à long terme du tissu industriel conduisant à la notion "d'entreprise virtuelle" ou "multi-centrique" (Saïdi-Kabeche Juillet 96) agrégeant de multiples sous-traitants autour d'un donneur d'ordres. Cela est également cohérent avec les approches actuelles de type "multi-agents". Deux principaux axes de recherche se dégagent, qui devraient permettre de mieux préciser et d'éclairer les problématiques de gestion de production dans le secteur des production multi-centriques. Ils concernent

- le problème crucial lié aux partages de compétences et de zones d'autonomie entre centres de décision ayant à coopérer autour d'objectifs communs, chacun ayant sa rationalité d'action spécifique, et tous étant liés par des "contrats" (au sens large du terme) et des règles (explicites ou implicites) d'évaluation de leurs décisions et de leurs actions.
- les enjeux de la centralisation/décentralisation des fonctions de décision en gestion de production et la préconisation d'un ensemble de stratégies et de tactiques de pilotage originales permettant d'atteindre la performance globale du système de production.

7. Bibliographie

- [1] Ayadi K., Arekion J, Kieffer J.P., Baudoin M. (1995), « A simulation approach for the scheduling of a complex manufacturing system, *1995 Summer Computer Simulation Conference*, Ottawa, July 23,26, 1995
- [2] Baillet P.(1994), Contribution à l'amélioration de la réactivité dans les systèmes de production, notamment par la mise en oeuvre des concepts de décentralisation des fonctions de décision, Thèse de Doctorat en Sciences, Université d'Aix-Marseille 3, nov. 1994
- [3] Dayoff J. E. and Atherton. R. W. (1984),"Simulation of VLSI manufacturing areas. *VLSI Design*, Dec. 84
- [4] Dufrene D., Gharbi A., Kieffer J.P., Rebouché J.Y., Villeneuve L. (1994), A discrete event simulation methodology for the design of flexible manufacturing systems, *1994 Summer Computer Simulation Conference*, La Jolla, San Diego, California, July 18-20 1994
- [5] Dufrene D., Kieffer J.P., Rebouché J.Y. (1994), « A method to improve the efficiency of the discrete event simulation projects, » *27th ISATA International Conference*, Aachen, Germany, Oct. 31st, Nov. 4th., 1994
- [6] Ferber J.(1989), Objets et agents : une étude des structures de représentation et de communication en intelligence artificielle, Thèse d'Etat, Université Paris 6, juin 1989
- [7] Glassey C.R. and Resende. M.G.C. (1988), « Closed- loop job release control for VLSI circuit manufacturing » , *IEEE Trans. Semic. Manuf.*, Vol 1, No.1. Feb 1988
- [8] Lawton J. W. et al.. (1990), « Workload regulating wafer release in a GaAS Fab facility » , *IEEE Intl Semi. Manuf. Science Symposium 1990*
- [9] Lozinski C , Glassey R (1988), « Bottleneck starvation indicators for shop floor control » , *IEEE Trans. Semi Manuf.*, Vol 1, No.4. Nov 1988
- [10] Panwalkar SS., Islander W. (1977), A survey of scheduling rules, *Operations Research.*, Vol 25 , 1. Jan-Feb 1977
- [11] Resende M (1986), Computer simulation of semiconductor Wafer fabrication, *note OCR 86-14.*, Sept 1986
- [12] Saidi - Kabeche Djoudja. Thèse de Doctorat. Exploration méthodologique des problèmes de pilotage de la production dans les systèmes de productifs " multi-centriques " Ecole Nationale Supérieure des Mines de Paris. Juillet 1996.
- [13] Steve c, H. Lu, Deepa Ramaswamy, P.R Kumar, Efficient Scheduling Policies to Reduce Mean and Variance of Cycle-Time in Semiconductor Manufacturing Plants, *IEEE Trans. Semi. Manuf.*, Vol 7 , No. 3. Aug 1994
- [14] Wien. L. M. (1988), « Scheduling semiconductor wafer fabrication » , *IEEE Trans. Vol.1.*, pp. 115-130 Aug. 1988

