

THESE

présentée par

Abderezak Touzene

pour obtenir le grade de **DOCTEUR**

de l'**INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE**

(Arrêté ministériel du 23 Novembre 1988)

Spécialité : INFORMATIQUE

=====

**RESOLUTION DES MODELES MARKOVIENS
SUR
MACHINES A MEMOIRES DISTRIBUEES**

=====

Date de soutenance : 21 Septembre 1992

Composition de jury :	D. Trystram	Président
	B. Plateau	Directeur de la thèse
	G. Bernard	Rapporteur
	J. C. Konig	Rapporteur
	J. M. Vincent	Examineur

Thèse préparée au sein du **Laboratoire de Génie Informatique.**

A mon père,

à ma mère,

A mes frères et sœurs,

REMERCIEMENTS

Tout d'abord, je tiens à remercier Brigitte Plateau qui, avec ses conseils ainsi que ses critiques constructives, a bien dirigé mon travail. Cette thèse doit beaucoup à son esprit de rigueur, à son entière disponibilité et surtout à sa patience.

Je remercie Denis Trystram pour l'honneur qu'il me fait en présidant le jury de cette thèse.

Mes remerciements vont également à Guy Bernard et Jean-Claude Konig pour leur acceptation d'être rapporteurs de cette thèse.

Je remercie Jean Marc Vincent qui a accepté de participer au jury.

Que tous les membres du laboratoire LGI et ceux du laboratoire LMC qui ont participé de près ou de loin à la réalisation de ce travail soient vivement remerciés.

J'adresse mes remerciements à Claire Dicrescenzo et Joelle Prevost de l'équipe technique du laboratoire LMC pour leur aide omniprésente et très appréciable.

Que tous mes amis et collègues qui ont rendus mon séjour en France agréable soient sincèrement remerciés.

Enfin, je tiens à exprimer toute ma reconnaissance aux différentes personnes qui ont participé, depuis mon enfance, dans ma formation intellectuelle et scientifique.

Table des matières

1	Présentation de la machine (MIMD) cible MEGANODE	5
1.1	Introduction	5
1.2	Description du MEGANODE à 128 Transputers	6
1.2.1	Vue générale	6
1.2.2	Réseau d'interconnexion pour Transputers	7
1.3	Modèle de communication pour le MEGANODE	9
1.3.1	Communication à l'intérieur d'un même tandem-node	9
1.3.2	Communication entre deux Transputers appartenant à deux tandem-nodes différents	12
1.3.3	Communication bi-directionnelle	13
1.3.4	Communication sur les liens en parallèle	15
1.4	Performance de l'unité de calcul du Transputer (T800/T414)	16
1.5	Taux de recouvrement des communications par du calcul	17
1.6	Conclusion	19
2	Communications dans les machines à mémoires distribuées	21
2.1	Introduction	21
2.2	Principaux facteurs qui influent sur le coût de communication	22

2.2.1	Mode de communication	23
2.2.1.1	Communication par commutation de message	23
2.2.1.2	Communication par commutation de circuit	23
2.2.1.3	Comparaison du coût de communication les différents modes	24
2.2.2	Topologie du réseau de communication	26
2.2.3	Type de communications intensive	26
2.2.3.1	Diffusion et multidiffusion (échange total)	27
2.2.3.2	Accumulation et échange total avec accumulation	27
2.2.3.3	Distribution, rassemblement et échange total personnalisé	27
2.2.3.4	Dualité entre les types de communication intensives	28
2.3	Echange total et échange total avec accumulation sur l'anneau	30
2.3.1	Echange total sur l'anneau	30
2.3.2	Echange total avec accumulation sur l'anneau	31
2.4	Echange total sur une topologie torique	33
2.4.1	Procédure d'échange total sur un tore ($\sqrt{p} \times \sqrt{p}$) avec \sqrt{p} impair	34
2.4.1.1	Notation	34
2.4.1.2	Position du problème	35
2.4.1.3	Exemple d'arbre de recouvrement F_{NB} du tore ($\sqrt{p} \times \sqrt{p}$) avec \sqrt{p} impair	36
2.4.1.4	Procédure d'échange total optimal pour le tore ($\sqrt{p} \times \sqrt{p}$) avec \sqrt{p} impair	40
2.4.2	Procédure d'échange total sur un tore ($\sqrt{p} \times \sqrt{p}$) avec \sqrt{p} pair	42
2.4.2.1	Exemple d'arbre de recouvrement F_{NB}^E pour un tore ($\sqrt{p} \times \sqrt{p}$) avec \sqrt{p} pair	43

Table des matières

2.4.2.2	Procédure d'échange total sur un tore ($\sqrt{p} \times \sqrt{p}$) avec \sqrt{p} pair	46
2.4.3	Implémentation de l'échange total	48
2.4.4	Conclusion sur l'échange total sur le tore	51
2.5	Echange total avec accumulation sur le tore	51
2.5.1	Principe de l'échange total avec accumulation sur le tore	51
2.5.2	Procédure d'accumulation en utilisant les arbres de recouvrement	51
2.5.3	Procédure d'échange total avec accumulation dans le tore	52
2.5.4	Implémentation de l'échange total avec accumulation	53
2.6	Procédure de communication échange partiel personnalisé sur le tore	55
2.6.1	Principe de l'échange partiel personnalisé sur le tore	55
2.6.2	Procédure de l'échange partiel personnalisé	56
2.6.3	Remarques sur la procédure de l'échange partiel personnalisé	57
2.7	Conclusion	57
3	Le parallélisme au service de l'évaluation des performances	59
3.1	Introduction	59
3.2	Méthode d'évaluation des performances et leur parallélisation	59
3.2.1	Méthode d'évaluation des performances utilisant directement le système réel	60
3.2.2	Méthode d'évaluation des performances utilisant des modèles du système réel	60
3.2.2.1	Méthodes de modélisation	60
3.2.2.2	Méthodes parallèles de résolution des modèles	61
3.3	Les méthodes numériques et leur parallélisation	62
3.3.1	Les méthodes généralistes	64

3.3.1.1	Méthode du type Jacobi	64
3.3.1.2	Méthode du type Gauss-Seidel	65
3.3.1.3	Comparaison entre les méthodes de type Jacobi et Gauss-Seidel	68
3.3.1.4	Méthodes itératives utilisant la technique de projection	68
3.3.2	Méthodes itératives spécialisés	70
3.3.2.1	Méthode itérative d'Agrégation-Désagrégation pour les processus (NCD)	71
3.3.2.2	Méthode de Neuts pour les processus (QBD)	75
3.4	Conclusion	76
4	Parallélisation du produit matrice vecteur	77
4.1	Introduction	77
4.2	Présentation des différentes implantation du produit matrice vecteur	78
4.2.1	Méthode de décomposition par colonne (C)	79
4.2.2	Méthode de décomposition par ligne (L)	80
4.2.3	Méthode de décomposition ligne-colonne (LC)	81
4.3	Comparaison théorique entre les trois méthodes	84
4.4	Expérimentation du produit matrice vecteur sur le MEGANODE	85
4.4.1	Résultat d'accélération des deux méthodes (C) et (LC)	85
4.4.2	Temps de calcul et communication des méthodes (C) et (LC)	86
4.5	Conclusion	88
5	Méthodes itératives parallèles de résolution des problèmes Markoviens	89
5.1	Introduction	89
5.2	Schéma itératif de base	91

Table des matières

5.3	Schéma itératif avec effet de Gauss-Seidel local	91
5.4	Schéma itératif à retard k	92
5.5	Amélioration du coût de calcul du produit pour un schéma à retard k	93
5.6	Schéma asynchrone libre version 1	94
5.7	Schéma asynchrone libre version 2	95
5.8	Résumé des schémas itératifs	96
5.9	Résultats expérimentaux sur les méthodes asynchrone	97
5.9.1	Plan d'expérience et choix des problèmes à tester	97
5.9.2	Matrices creuses	97
5.9.3	Matrices du type presque décomposable (NCD)	98
5.9.4	Matrices tridiagonales par bloc	98
5.9.5	Matrices du type marche aléatoire sur une grille 2D	98
5.9.6	Conclusion	98
5.9.7	Figures	99
6	Parallélisation du produit vecteur par le descripteur d'un RAS	107
6.1	Introduction	107
6.2	Principe de la multiplication vecteur par le descripteur d'un (RAS)	108
6.2.1	Ordre des composantes dans le vecteur π	109
6.2.2	Définitions des mélanges parfaits	110
6.2.3	Mélange parfait des composantes dans une base de dimension 2	111
6.2.4	Mélange parfait des composantes dans une base de dimension 3	120
6.2.5	Calcul de $\pi [\otimes_{i=1}^N M_{ti}]$ en utilisant les mélanges parfaits dans une base de dimension 2	122

6.2.6	Calcul de $\pi [\otimes_{i=1}^N M_{ti}]$ en utilisant les mélanges parfaits dans une base de dimension 3	124
6.3	Présentation des différentes approches de parallélisation	130
6.4	Première approche : parallélisation du calcul d'un terme ($\pi [\otimes_{i=1}^N M_{ti}]$)	131
6.4.1	Introduction	131
6.4.2	Parallélisation utilisant les mélanges parfaits dans une base de dimension 2	131
6.4.2.1	Problème de placement pour résoudre les mélanges parfaits dans une base de dimension 2	132
6.4.2.2	Détails de parallélisation de $\pi \prod_{i=1}^N [S_{b_i B_i} (Id_{B_i} \otimes M_i)]$	138
6.4.3	Parallélisation utilisant les mélanges parfaits dans une base de dimension 3	144
6.4.4	Parallélisation <i>mixte</i> combinant les mélanges parfaits dans une base respectivement de dimension 2 et 3	146
6.4.4.1	Impact du choix du niveau k sur la recherche du bon grain de parallélisme	146
6.4.4.2	Evaluation du coût d'une affectation	147
6.4.4.3	Critère de choix de l'ordre lexicographique	148
6.4.4.4	Algorithme de la méthode mixte:	154
6.5	Deuxième approche : parallélisation par distribution des termes de la somme $\pi [\sum_{t=1}^C (\otimes_{i=1}^N M_{ti})]$	158
6.6	Troisième approche : parallélisation de la somme et des termes de la somme	160
6.6.1	Réseaux toriques pour effectuer les termes et la somme	161
6.6.2	Réseaux semi-toriques pour effectuer les termes et réseaux anneaux pour la somme	162
6.6.3	Réseaux anneaux pour effectuer les termes et réseaux anneaux pour la somme	165

Table des matières

6.6.4	Comparaison entre les trois stratégies et choix de la meilleure solution	165
6.6.5	Etude de l'accélération pour la solution choisie	167
6.7	Comparaison entre les 3 approches de parallélisation et choix de la plus appropriée	168
6.8	Amélioration et expérimentation de l'approche choisie	168
6.8.1	Amélioration de la méthode en utilisant les propriétés des mélanges parfaits	169
6.8.1.1	Comparaison expérimentale entre le mélange amélioré et le mélange parfait naif	171
6.8.1.2	Comparaison expérimentale entre le temps du mélange parfait et le temps du produit correspondant	172
6.8.1.3	Performances de cette approche de parallélisation	172
6.8.2	Amélioration de la méthode en utilisant le recouvrement des communications par les calculs	175
6.8.2.1	Premier cas:le temps des produits locaux est du même ordre que le temps d'un mélange parfait externe	176
6.8.2.2	Deuxième cas : Le temps des produits locaux est supérieur au temps total de tous les produits externes	180
6.8.2.3	Algorithme général utilisant la technique de recouvrement	182
6.9	Conclusion	183
A	Modèle de calcul asynchrone	185
A.1	Asynchronisme total	186
A.2	Application au système d'équations linéaires	188
A.3	Modèle de calcul asynchrone partiel	189
A.4	Application au calcul du vecteur stationnaire de probabilités	191

Table des matières

B	Terminaison des algorithmes itératifs parallèles	193
B.1	Convergence globale	193
B.2	Convergence locale	195
C	Expérimentation des méthodes de puissances classique et réactualisée en séquentiel	197
C.1	Plan d'expérience	197
C.2	Matrices creuses	198
C.3	Matrices du type presque décomposable (NCD)	198
C.4	Matrices triangulaires par bloc	198
D	Génération des matrices	203
D.1	Matrices creuses	203
D.2	Matrices du type presque décomposable (NCD)	203
D.3	Matrices du type tridiagonales blocs	203
D.4	Matrice de transition de la marche aléatoire sur une grille 2D	204
E	Algorithmes de placement	205
E.1	Algorithme de Bokhari "échange de pairs"	205
E.2	Algorithme de recuit simulé	206
F	Résultats expérimentaux sur le choix de l'ordre lexicographique	209
G	Recouvrement des communications par les calculs	221