Utilisation de PM^2 pour la programmation data-parallèle sur réseau de stations :
le modèle des réseaux associatifs et la bibliothèque ANET

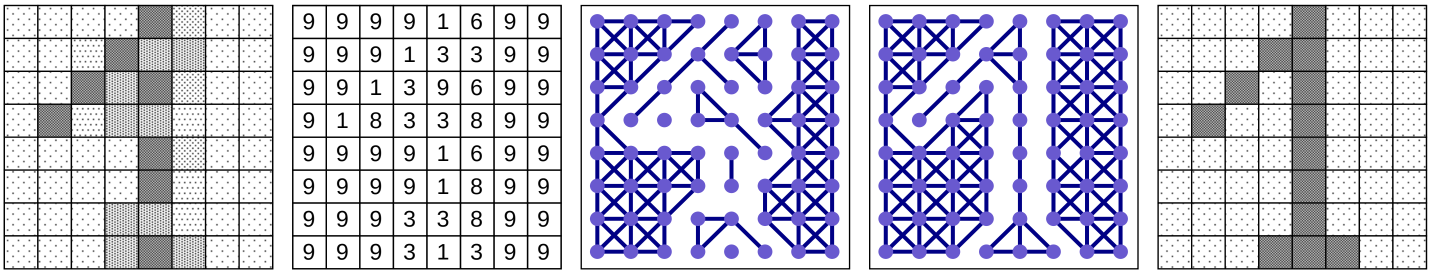
Bertrand Ducourthial
Laboratoire de Recherche en Informatique
Institut d'Électronique Fondamentale
Université Paris-Sud FRANCE
`ducourth@lri.fr`
`www.ief.u-psud.fr/~ducourth`

5 mars 1999

Journées IhPerf sur les processus légers
Grenoble 4 et 5 mars 1999

Origine de l'étude

- Modèle des réseaux associatifs [Mérigot 1992]
  - traitement d'images et architectures spécialisées
  - image = pixels + relations inter-pixels
    - data-parallélisme $\rightsquigarrow$ image : variable parallèle (*pvar*)
    - représentation des “objets” d'une image par un sous-réseau, dans un réseau de base fixé (*graphe maille*)
  - primitives de calcul global sur les sous-réseaux
  - Maille Associative : multi-processeurs SIMD



9	9	9	9	1	6	9	9
9	9	9	1	3	3	9	9
9	9	1	3	9	6	9	9
9	1	8	3	3	8	9	9
9	9	9	9	1	6	9	9
9	9	9	9	1	8	9	9
9	9	9	3	3	8	9	9
9	9	9	3	1	3	9	9

- Thèse
  - extension du modèle : primitive asynchrone
  - nouvelles applications : image, graphe, géométrie, ...
  - mise en œuvre du modèle sur réseau de stations

- Intérêt
  - simplifier la programmation d'applications complexes $e.g.$, segmentations d'images
  - environnement logiciel sur architecture classique

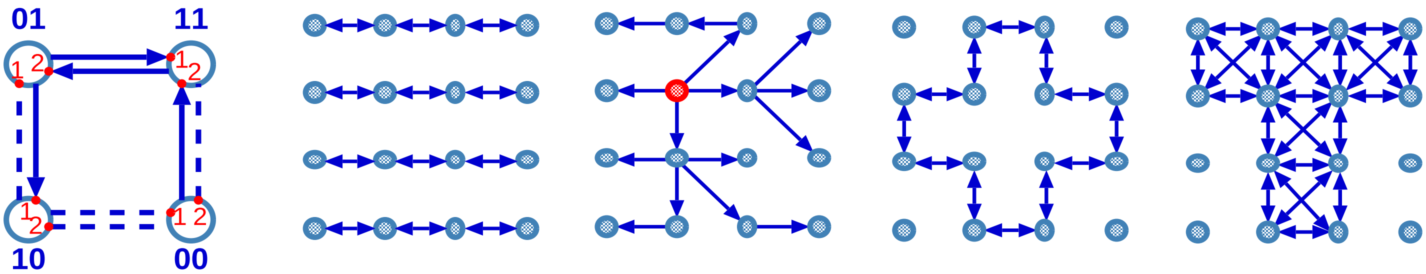
Plan

- Modèle
  - codage des objets
  - primitive `direct-association`
  - exemple : composantes fortement connexes
  - conditions assurant la définition :
    graphes préfixes et r-opérateurs
  - du modèle à l'algorithmique
- Algorithmique
  - image : distance
  - géométrie : domaines de Voronoï
  - segmentation d'images en régions de Voronoï
- Mise en œuvre : ANET
  - objectifs
  - structure
  - utilisation de PM^2
  - parallélisation des primitives
  - test : r-opérateurs et grain du calcul
  - test : asynchronisme
- Conclusion

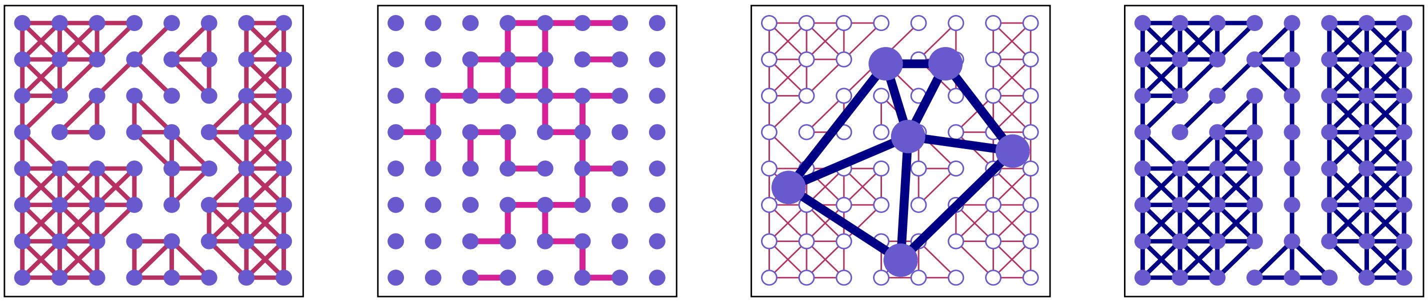
Codage des objets

Technique des masques

- masques sur les arcs entrants dans un réseau de base
- $\rightsquigarrow$ pvar de masques



Niveau d'abstraction supplémentaire : graphe virtuel



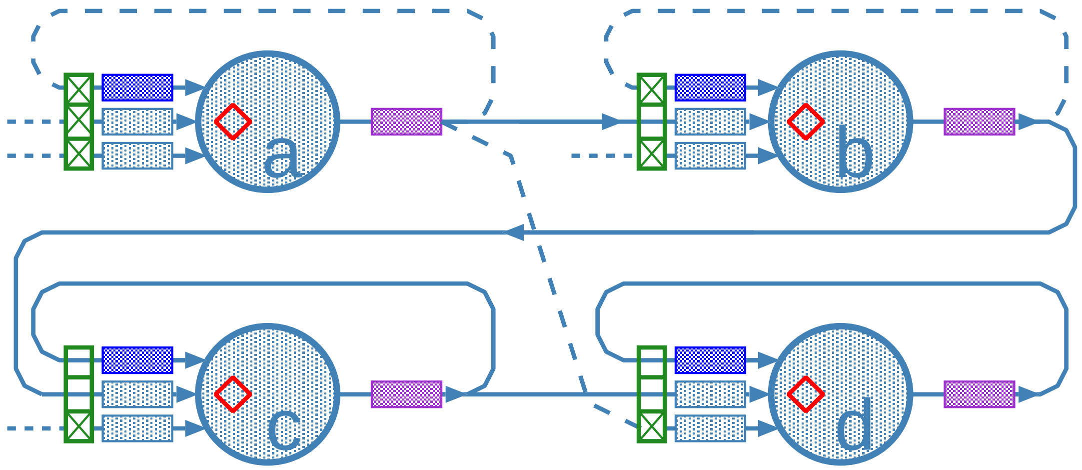
- graphe des pixels $\rightsquigarrow$ graphe des régions
- traitements de haut niveau, sur les objets de l'image
- codage par deux variables parallèles (*pvars de masques*)

Intérêt

- objets irréguliers dans une machine virtuelle d'instanciation
- souplesse d'utilisation ; gains mémoire
- possibilités illimitées (plongements)

Primitive `direct-association`

- Réseau codé par les masques ; entités calculantes

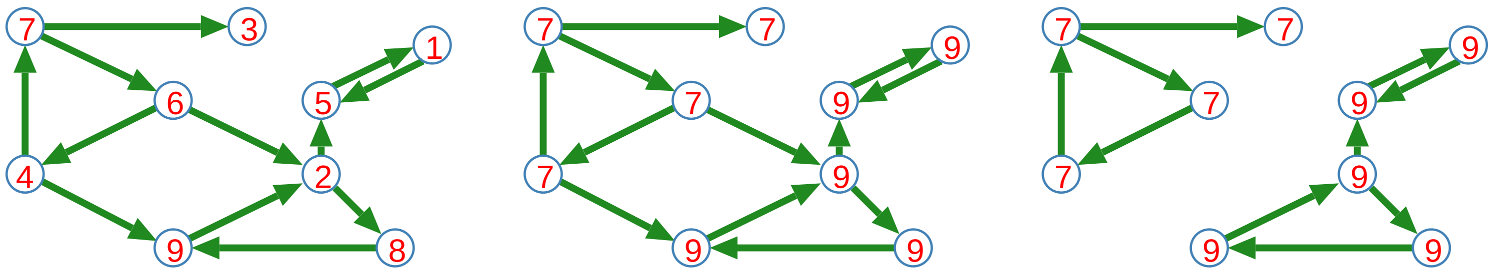


- Calcul local avec un opérateur $\diamond$ binaire ou n-aire

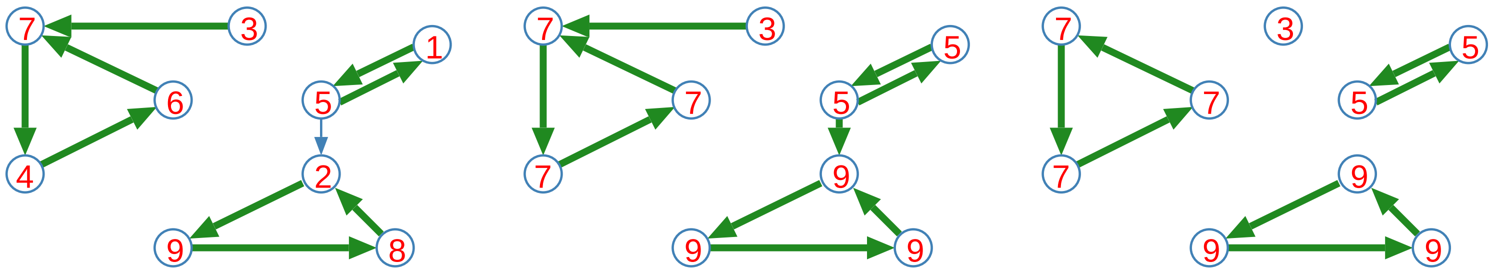
$$x_{\text{out}} \leftarrow x_0 \diamond \cdots \diamond x_{\delta_v^-} \qquad x_{\text{out}} \leftarrow \diamond \left(x_0, \ldots, x_{\delta_v^-} \right)$$

- Calcul global par coopération de calculs locaux
  - $q \leftarrow \diamond$-*direct-association*(G, p)
  - calculs locaux non coordonnés
  - fin du calcul : aucune nouvelle valeur n'est produite $\rightsquigarrow$ stabilisation
  - défini si terminaison et déterminisme

Exemple : composantes fortement connexes



a. Graphe initial, pvar index.   b. max-direct-association   c. link-with-equals



d. Graphe inversé, pvar index.   e. max-direct-association   f. link-with-equals

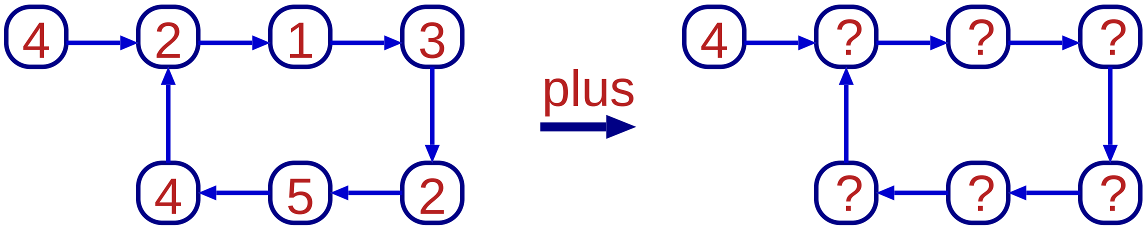
Algorithme 1 CFC(**pvar**⟨ **masque** ⟩ G)

pvar⟨ **masque** ⟩ G_{tmp}
pvar p
répéter
  $G_{tmp} \leftarrow G$
  $p \leftarrow$ `max-direct-association(`G`,`*index*`)`
  $G \leftarrow$ `link-with-equals(`G, p, p`)`
  $p \leftarrow$ `max-direct-association(`G^{-1}`,`*index*`)`
  $G \leftarrow$ `link-with-equals(`G, p, p`)`
jusqu'à ce que $G \neq G_{tmp}$
retourner G

Conditions assurant la définition (1)

direct-association :

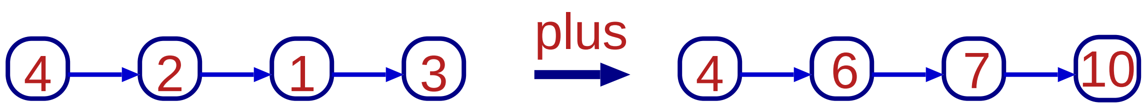
- peu de contraintes $\rightsquigarrow$ facilement mise en œuvre
  mais quels opérateurs sur quels réseaux ?
- exemple : plus-direct-association



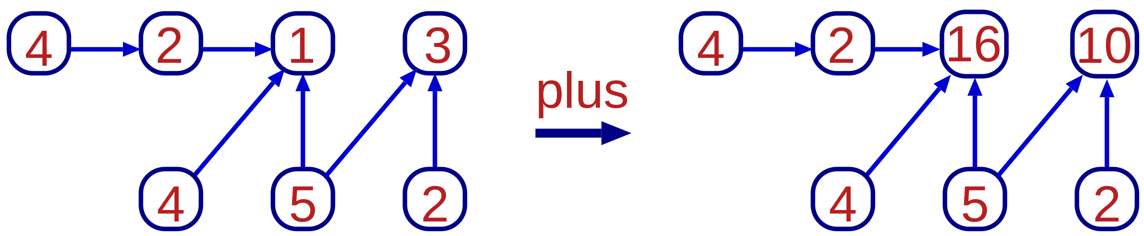
- $\rightsquigarrow$ *absence de circuit ou idempotence* $(x \diamond x = x)$

Absence de circuit :

- scan vector model [Blelloch 1990]

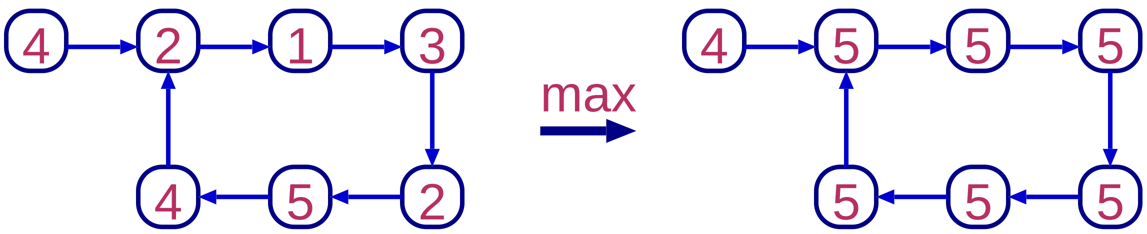


- généralisation aux *graphes préfixes*
  (connexions d'arborescences)

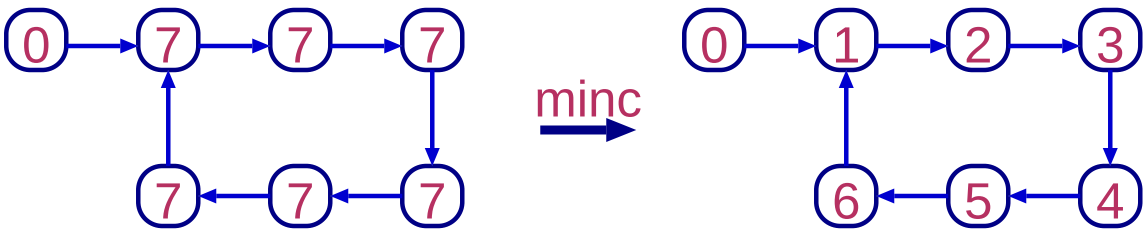


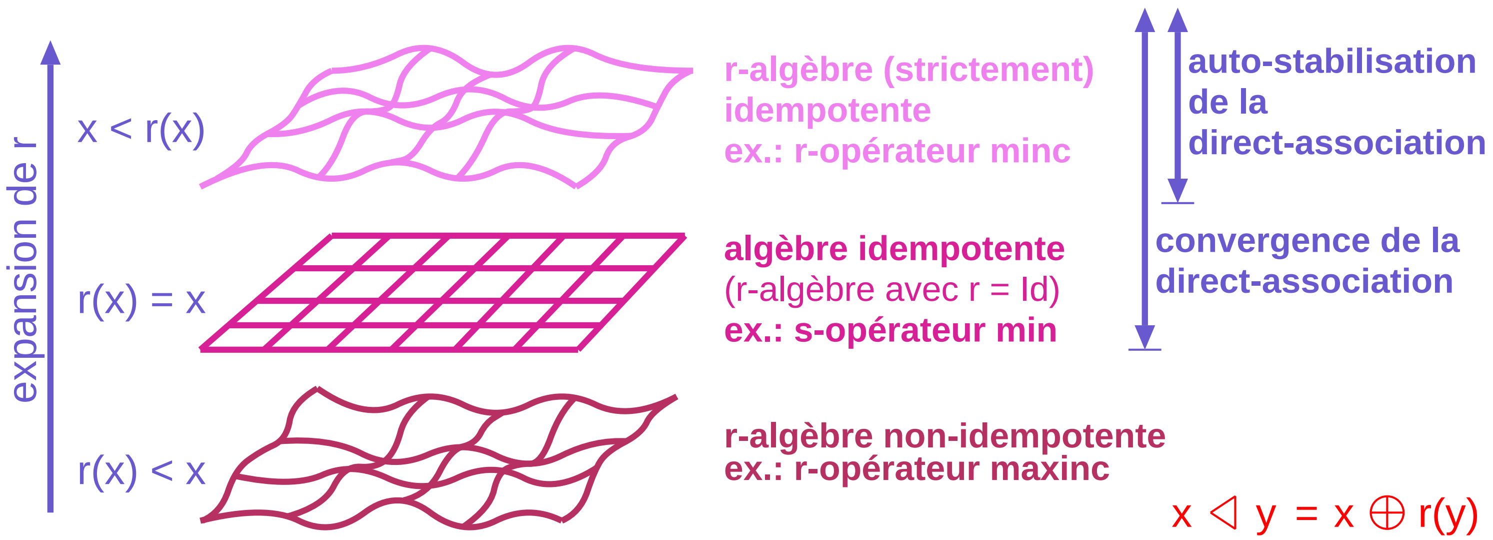
Conditions assurant la définition (2)

- Idempotence :
  - s-opérateur : associatif, commutatif et idempotent
    $\rightsquigarrow$ *total algorithms* (ou *waves*), et infimum [Tel 1991]



  - généralisation aux r-*opérateurs* idempotents
    exemple : minc$(x, y) =$ min$(x, y + 1)$





$x \lhd y = x \oplus r(y)$

Du modèle à l'algorithmique

Réseau

- sous-réseaux dynamiques d'un réseau de base
- maille : bonne machine virtuelle (plongements)
- abstraction avec les graphes virtuels

`◇-direct-association`(G, p)

- spécification des calculs locaux, et non du calcul global
- conditions sur le réseau G et/ou l'opérateur $\diamond$

réseau	opérateur
chaîne $\rightarrow$ graphe préfixe	opérateur quelconque
graphe quelconque	s-opérateur $\rightarrow$ r-opérateur

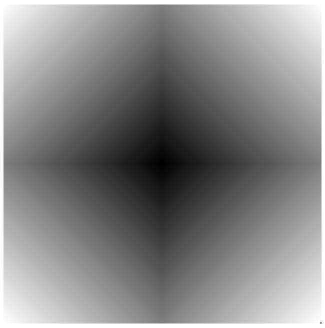
- lien calcul parallèle $\leftrightarrow$ calcul distribué
- utilisation du réseau sous-jacent et de l'asynchronisme

Applications

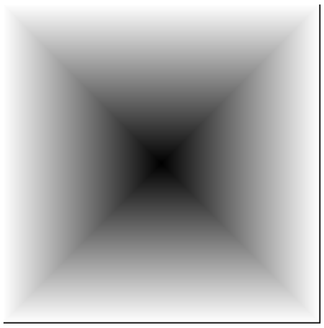
- graphe *tri topologique, CC, CFC, bfs, dfs, sssp, diamètre etc.*
- géométrie *polygones et <u>régions de Voronoï</u>, enveloppe convexe*
- image *squelette, contours, <u>distance</u>, morphologie, segmentation par ligne de montée de eaux, <u>segmentation de Voronoï</u>*

Image : distance

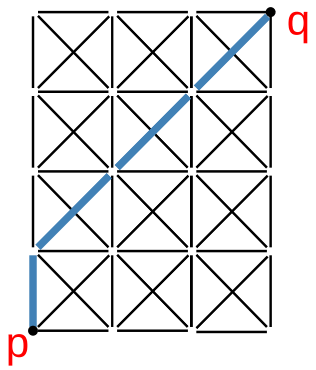
- Transformée en distance
  - $d_4(p,q) = |x_p - x_q| + |y_p - y_q|$
  - $d_8(p,q) = \max\left[|x_p - x_q|, |y_p - y_q|\right]$



d_4

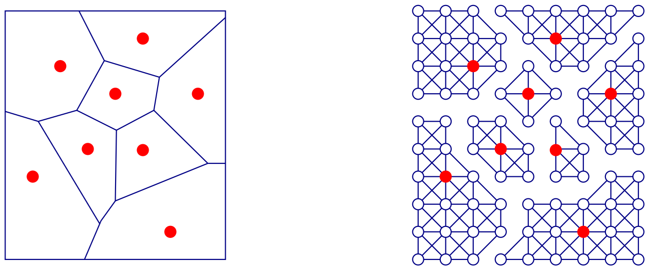


d_8

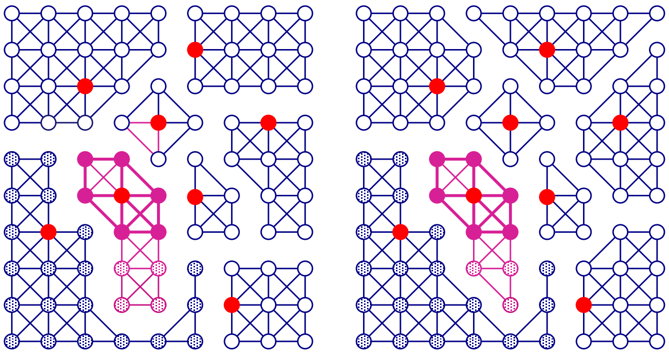


- minc-direct-association(*grille*, *image*)
  - $\text{minc}(x,y) = \min(x, y+1)$
  - un seul calcul asynchrone plus gros au lieu de plusieurs petits calculs synchrones
  - $\rightsquigarrow$ expressivité
    grossissement du grain du calcul

Géométrie : domaines de Voronoï

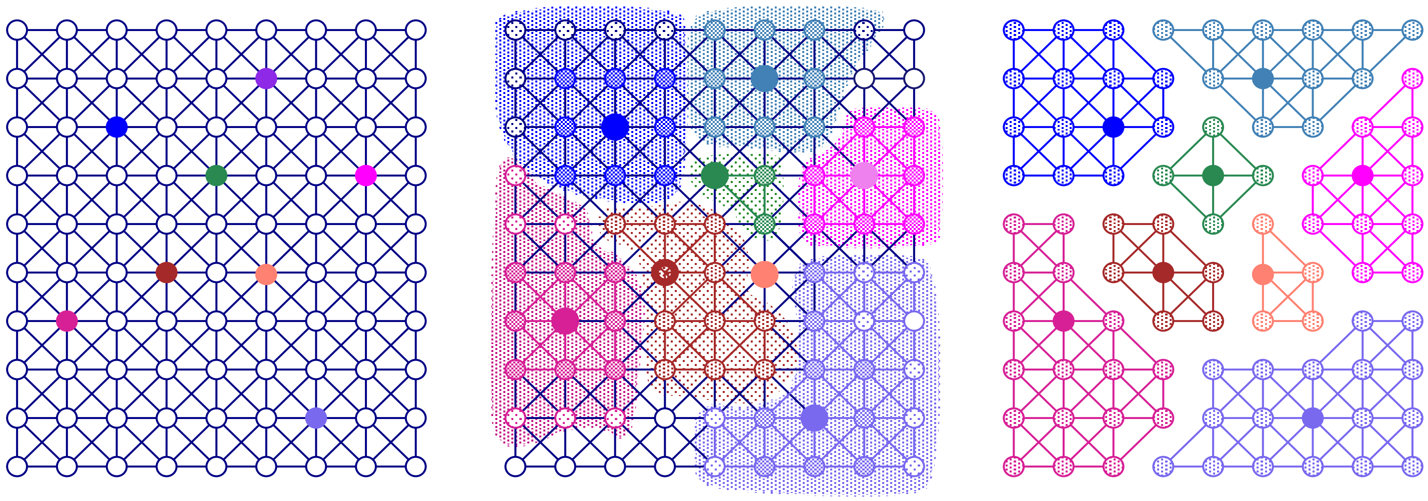


Calcul data-parallèle synchrone



régions initiales quelconques
tant qu'une région évolue, faire
  tant qu'une région voisine est à visiter, faire
    calcul de la distance à la graine voisine
    rendre les sommets de la graine voisine
  fin tant que
fin tant que

Calcul data-parallèle asynchrone

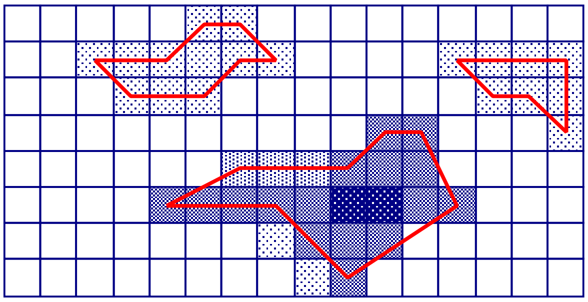
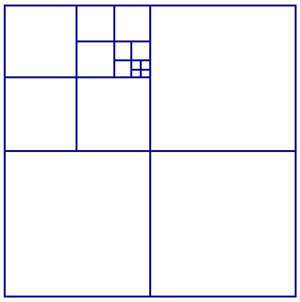
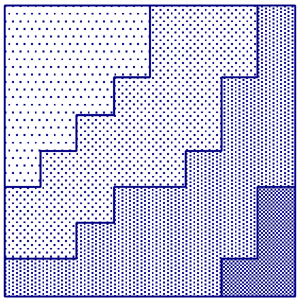


- une seule `minc-direct-association`(*maille*, p)
- $\rightsquigarrow$ suppression d'itérations

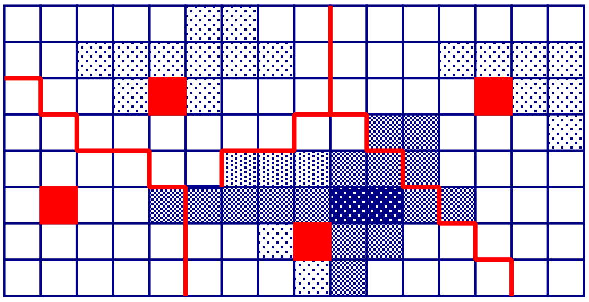
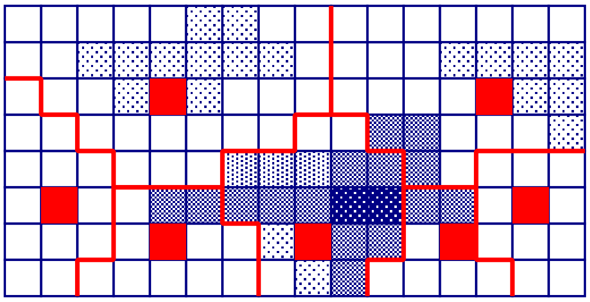
Segmentation d'images

Segmentation d'images

- déterminer les objets
- critères d'homogénéité, de texture, de forme...
- découpe et fusion [Horowitz et Pavlidis 1976]

Segmentation de Voronoï [Ahuja et Shachte, 1985]

- critères d'homogénéité et de forme $\rightsquigarrow$ complexe
- parallélisme de données difficile à exploiter
- réseaux associatifs :
  - $\rightsquigarrow$ découpe avec des minc-`direct-associations`, fusion à l'aide des graphes virtuels

ANET (1). Objectifs

Réseau associatif

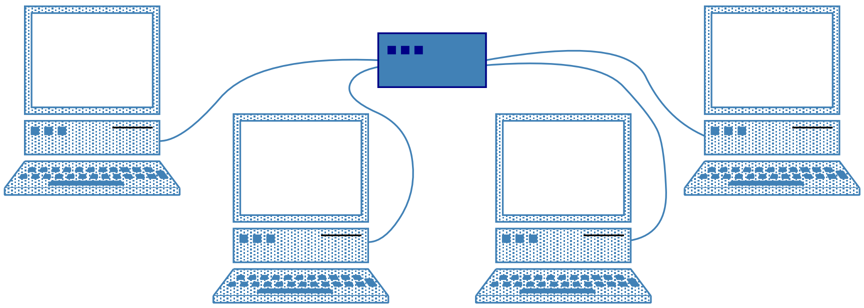
- réseau de base + sous-réseaux $\rightsquigarrow$ objets irréguliers
  exemple : régions de Voronoï
- primitives `associations`, calculs asynchrones
- irrégularité des traitements
  exemple : segmentation d'images

Paralléliser les primitives de calcul en respectant :

- leurs spécificités ($e.g.$, : asynchronisme)
- l'irrégularité de la forme des données
- l'évolution des formes très dynamique
- l'irrégularité des traitements

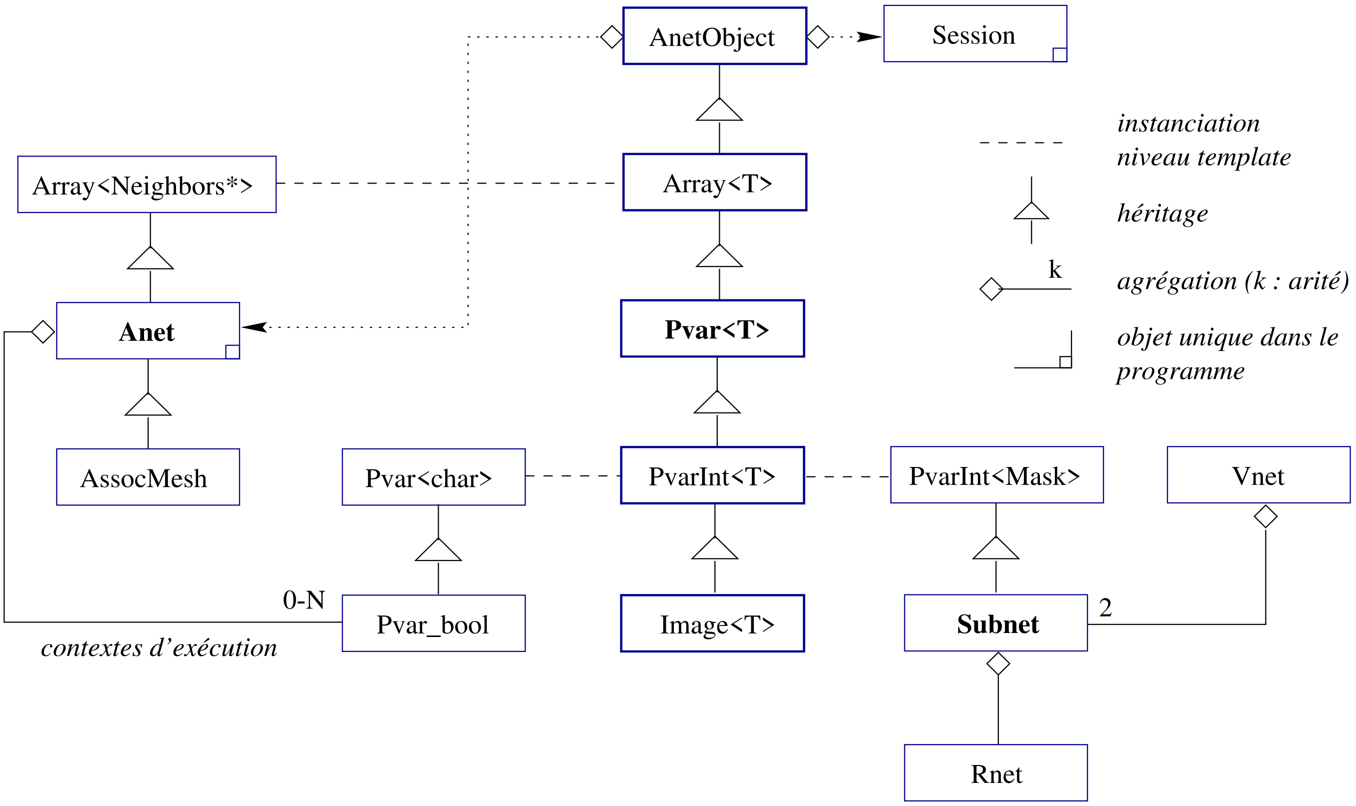
Plateformes visées

- PCs multi-processeurs
- grappes de PCs



ANET (2). Structure

Bibliothèque C++



Intérêt

définition d'objets par l'utilisateur,
et d'opérateurs pour les manipuler

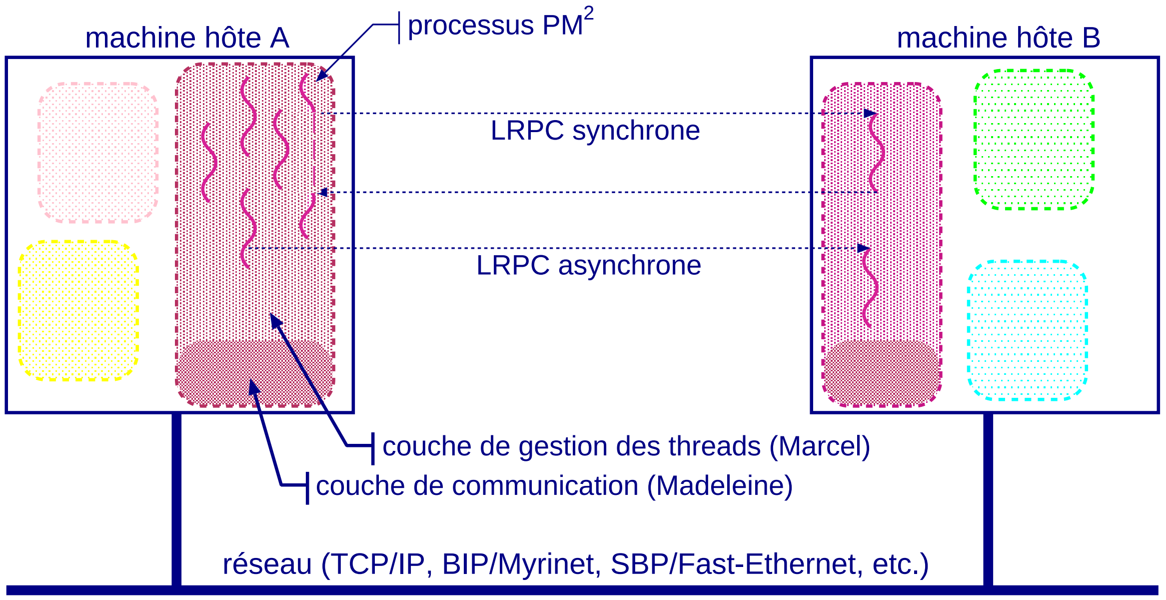
ANET (3). Utilisation de PM^2

Besoins

- placement souple sur les ressources de calcul
- gestion de l'irrégularité préventive
- plusieurs calculs, calculs et communications
- irrégularité $\rightsquigarrow$ équilibrage de charge (à terme)
- grain fin $\rightsquigarrow$ mécanismes peu chers

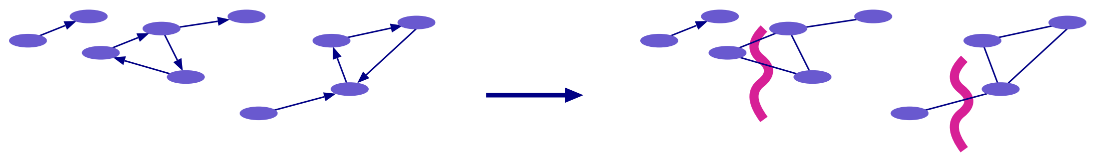
PM^2 : *Parallel Multithreaded Machine*

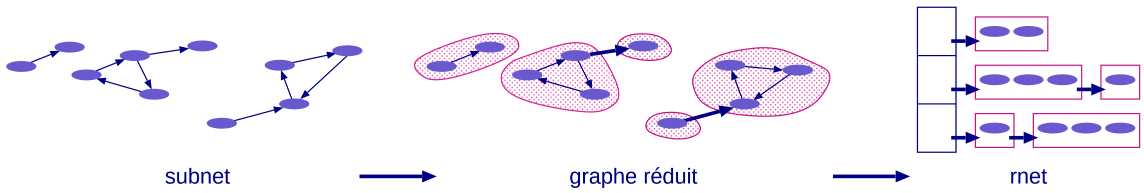
- processus “légers” $\rightsquigarrow$ dynamisme
- exécutions multi-fils
- recouvrement calcul / communication
- exécution à distance / migration
  $\rightsquigarrow$ équilibrage de charge



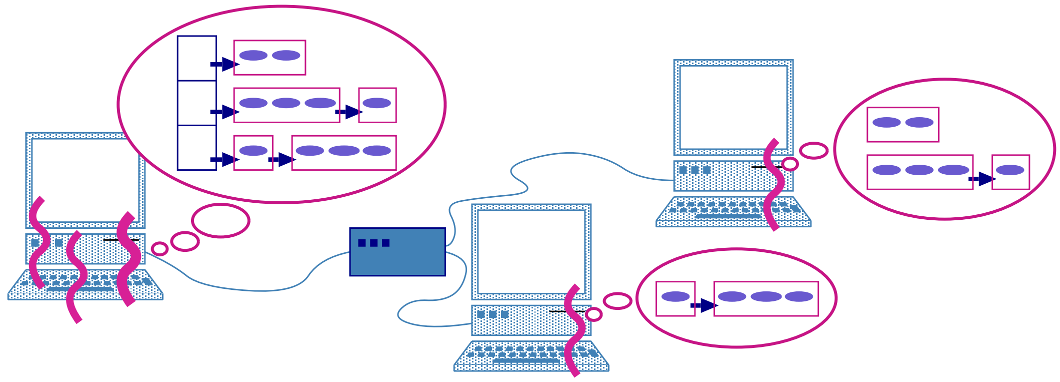
ANET (4). Parallélisation

- Mode maître-esclave
  - éclatement des objets peu envisageable :
    calculs globaux asynchrones sur les objets
    $\leadsto$ nombreuses “communications” dans les objets
  - distribution des objets peu rentable :
    segmentations $\leadsto$ formes très dynamiques
    $\leadsto$ redistributions très nombreuses
- Gestion de l'irrégularité des données





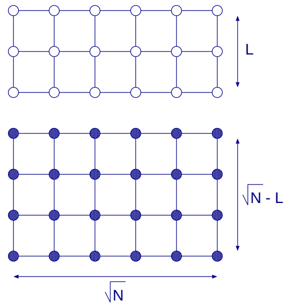
- Gestion de l'irrégularité des traitements



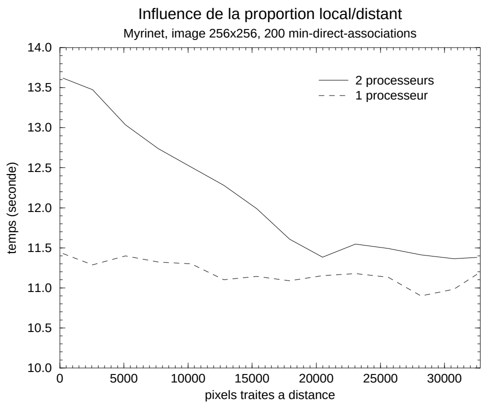
  - exécutions à distance $\leadsto$ ok si grain pas trop fin

Intérêt des r-opérateurs : grain du calcul

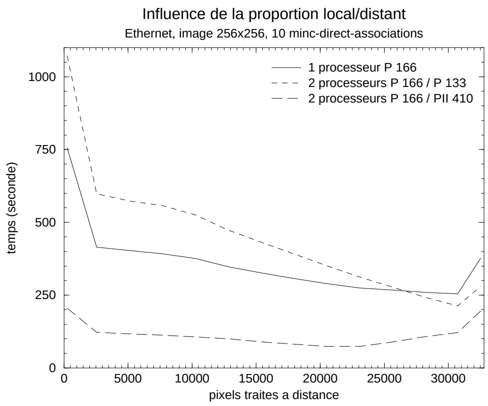
Comparaisons d'opérateurs



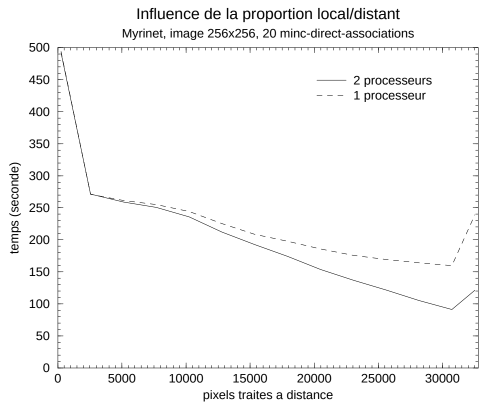
- deux processeurs
- un seul calcul global
- comparaison de min et minc
- influence réseau et machines



Myrinet min

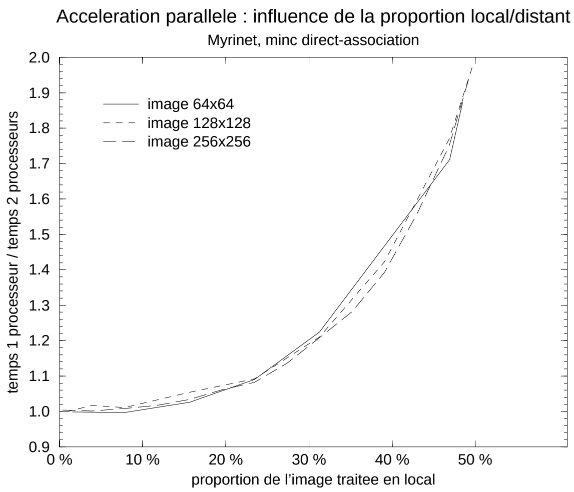


Ethernet minc



Myrinet minc

Interprétation



- r-opérateurs $\rightsquigarrow$ grain plus gros $\rightsquigarrow$ facilite la parallélisation
- intérêt des réseaux rapides
- influence de l'équilibrage de charge

Intérêt des r-opérateurs : asynchronisme

Découpe de Voronoï sur un seul processeur

Taille image	Type image	ANET s-opérateurs	ANET r-opérateurs	rapport s- / r-
Images 64×64 10 racines	forme	10,48	1,55	6,76
	icoron	6,98	1,44	4,84
	imuscle	8,77	1,16	7,56
	roks	9,78	1,23	7,95
	moyenne	*9,04*	*1,29*	*6,77*
Images 128×128 40 racines	forme	88,14	7,20	12,24
	icoron	41,05	5,53	7,42
	imuscle	52,28	5,47	9,55
	roks	57,64	5,92	9,73
	moyenne	*57,75*	*5,9*	*9,73*
Images 256×256 160 racines	forme	454,82	42,65	10,66
	icoron	230,40	28,96	7,95
	imuscle	340,77	31,74	10,73
	roks	301,50	30,25	9,96
	moyenne	*328,89*	*32,14*	*9,82*

Remplacement d'itérations

- s-opérateurs : $p \times q$ itérations $O(N)$
- r-opérateurs : p itérations en $O(N^2)$
- expériences : q, N, et α, $1 < \alpha \leq 2$, tels que

$$N^\alpha < q \times N$$

- N borné (images)

Conclusion

Modèle des réseaux associatifs

- modèle de programmation data-parallèle original
  masque : puissance du calcul sur réseau dynamique
  asynchronisme : algorithmique $\rightsquigarrow$ expressivité
  parallélisation $\rightsquigarrow$ efficacité
- lien entre scans et calculs d'infimum
  extensions avec graphes préfixes et r-algèbres
- algorithmique variée *image, graphe, géométrie, optimisation*

Mise en œuvre : ANET

- primitives des réseaux associatifs
- gestion de l'irrégularité, structure objet
- utilisation : UNIX, Myrinet (PM^2 + BIP), Ethernet
- réseaux associatifs $\rightsquigarrow$ placement de *threads*
  $\rightsquigarrow$ équilibrage de charge (en cours)

Perspectives

court terme :

- équilibrage de charge, optimisations diverses
- nouveaux tests, nouvelles applications
- plateforme Ethernet Gigabit de l'IEF
- SMP ?

moyen terme :

- serveur multi-application de traitement multimédia
  segmentation, compression, indexation, normes MPEG 4 et 7