

Equilibrage de charge et distribution de données pour des simulations parallèles irrégulières couplées

Maîtres de stage

Olivier Coulaud ; Aurélien Esnard ; Jean Roman

Tuteur

François PELLEGRINI

Période du stage

09/02/2009 – 30/09/2009

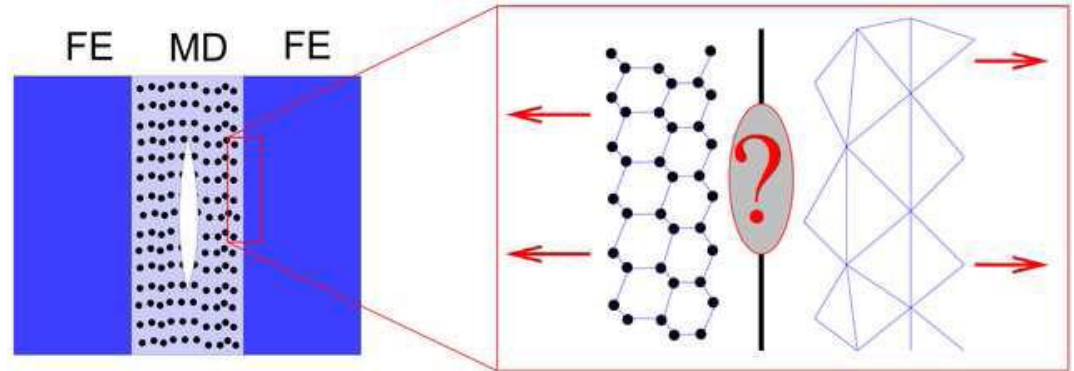
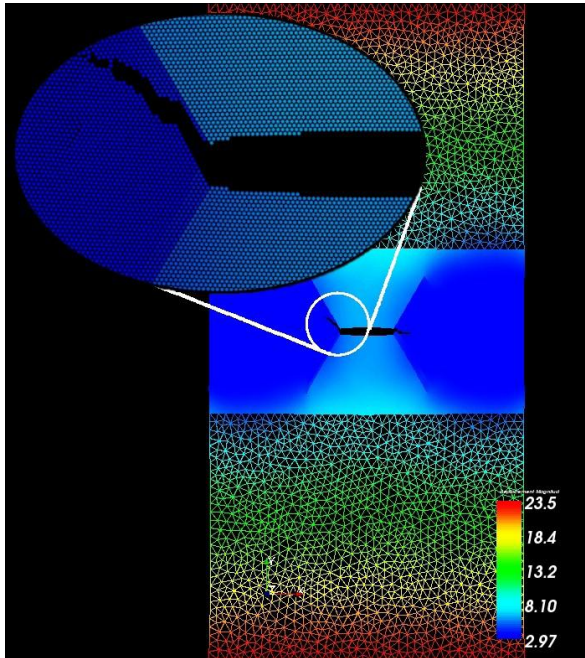
Projet de Fin d'Etudes - Master Recherche

Mohamed Amine EL AFRIT

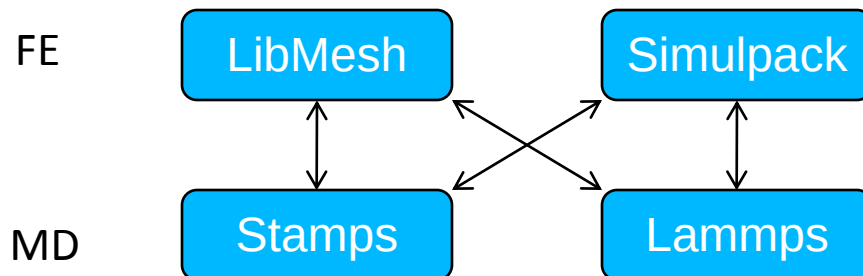
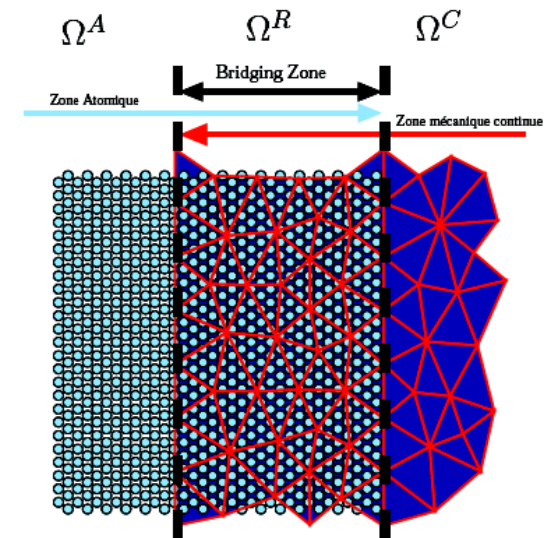
PLAN

- **Introduction**
- Modélisation
- Résultats
- Conclusion et perspectives

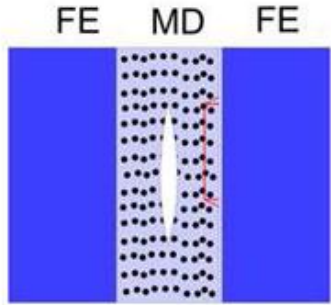
Cadre général



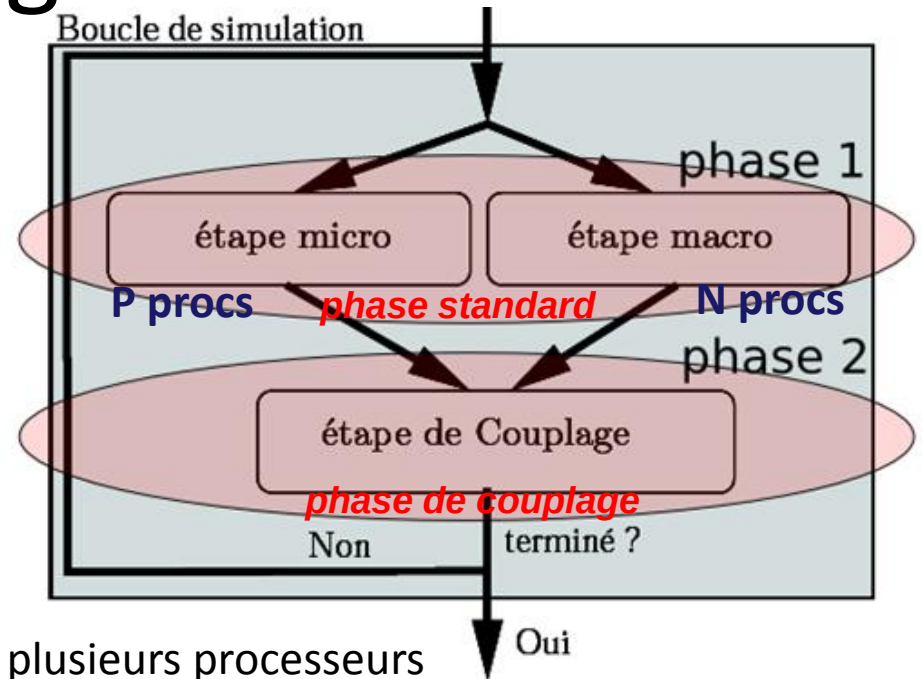
- Etude des propriétés des matériaux
- Couplage de code simulant à deux échelles différentes
- **LibMultiScale** : couplage générique de codes



Cadre général

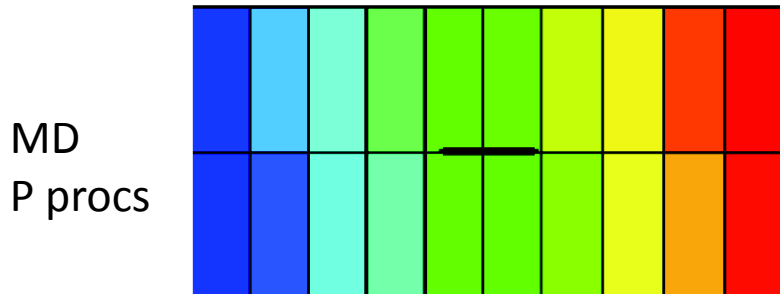


- *FE = zones continue*
- *MD = zones moléculaire*
- *Deux phases = phase standard & phase de couplage*



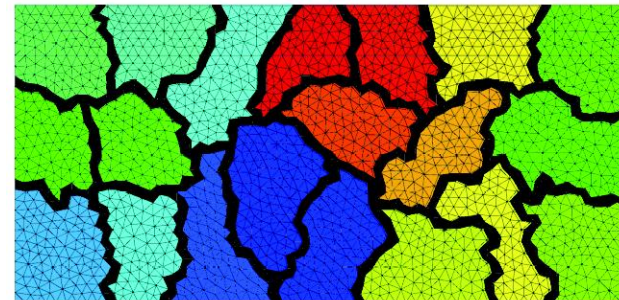
Une simulation parallèle sur plusieurs processeurs

Décomposition de la zone atomique (MD) en boîtes



Code MD // sur 20 procs

Décomposition de la zone continue (FE) par partitionnement de graphe



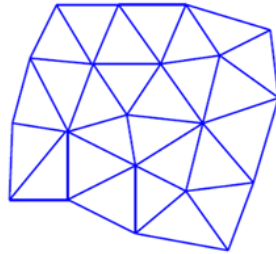
Code FE // sur 20 procs

couplé avec

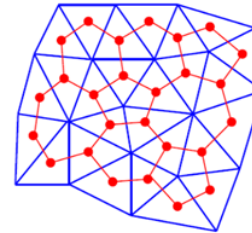
Problème de partitionnement

Partitionner un graphe = **diviser** un graphe en **N parties de même taille** (avec une tolérance) en **minimisant la coupe** (nombre d'arêtes entre les parties)

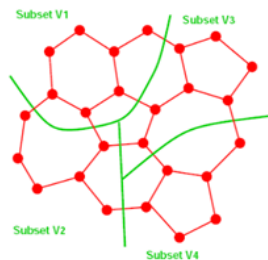
[15] *Teresco, Devine and Flaherty*



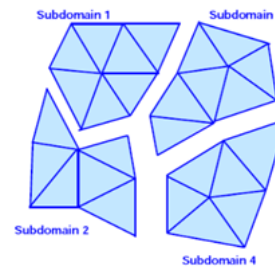
1- Maillage



2- Graphe du maillage

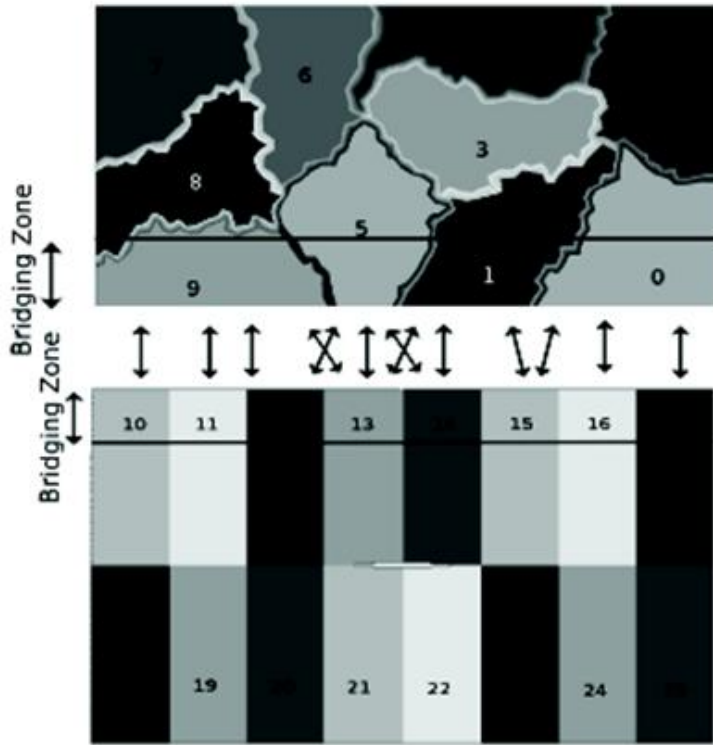


3- Partitionnement du graphe

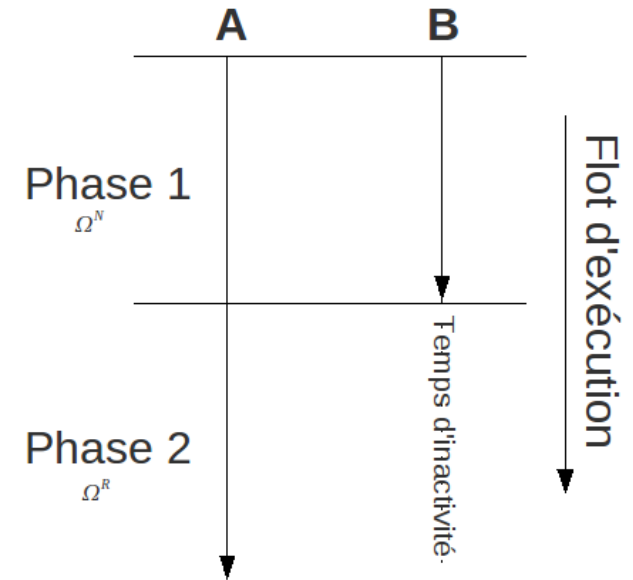
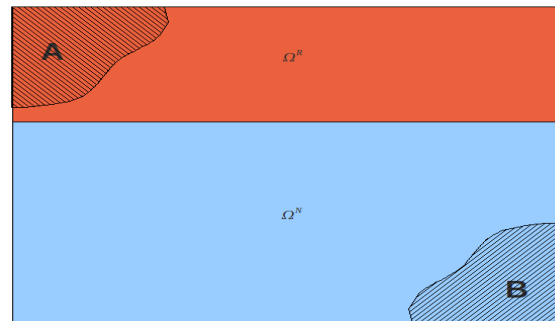


4- Distribution sur des proc

Problématique



- Mauvais équilibre entre les 2 phases
- Mauvais équilibre pendant la phase de couplage
- Schéma de communication FE↔MD mal équilibré



Problématique

Objectifs

1. Equilibrer la charge pendant les deux phases

deux phases = phase standard & phase de couplage

Problématique

Objectifs

1 - Equilibrer la charge pendant les deux phases

deux phases = phase standard & phase de couplage

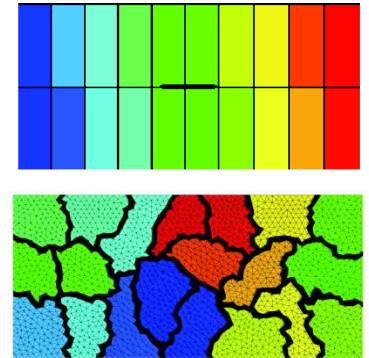
2 - Améliorer les communications FE-MD

FE = processeurs éléments finis

MD = processeurs dynamique moléculaire

Problématique

Objectifs



1 - Equilibrer la charge pendant les deux phases

deux phases = phase standard & phase de couplage

2 - Améliorer les communications FE-MD

FE = processeurs éléments finis

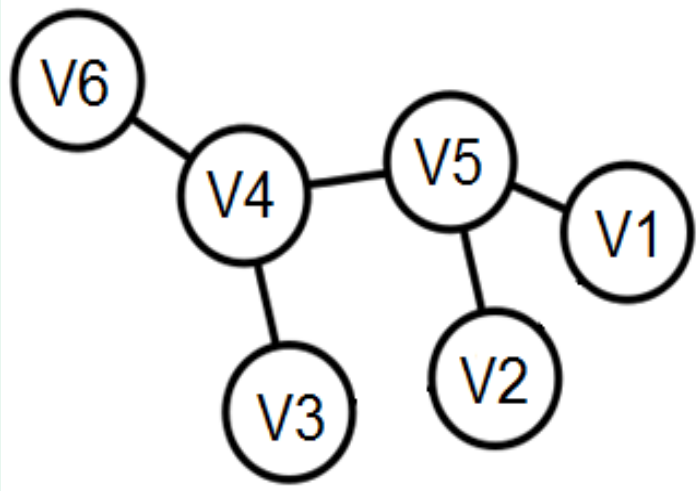
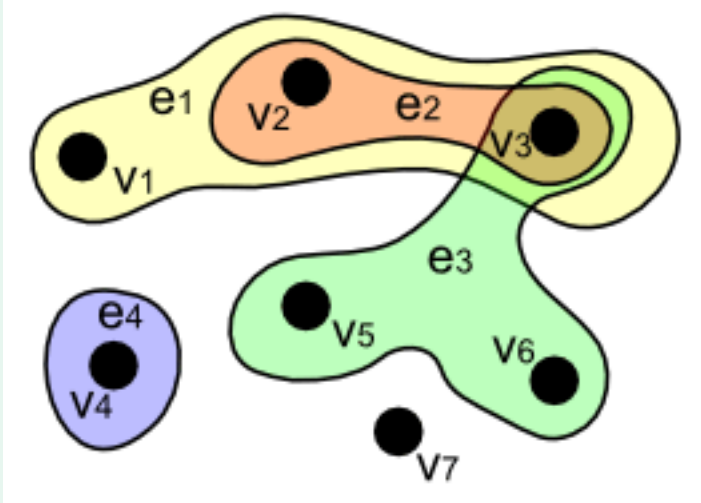
MD = processeurs dynamique moléculaire

**→ On se concentre sur le partitionnement des 2 phases du code FE
(en supposant que le partitionnement du code MD fixe)**

Notions préliminaires

Graphe vs Hypergraphe

l'Hypergraphe généralise la notion de Graphe. première fois par Claude Berge en 1960 [2]

Graphe	Hypergraphe
deux sommets par arête	plusieurs sommets par hyper-arête
 $V = \{v_1, v_2, v_3, v_4, v_5\}$ $E = \{e_1, e_2, e_3, e_4, e_5\}$ $= \{\{v_1, v_5\}, \{v_2, v_5\}, \{v_5, v_4\}, \{v_4, v_3\}, \{v_4, v_6\}\}$	 $V = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7\}$ $E = \{h_1, h_2, h_3, h_4\}$ $= \{\{v_1, v_2, v_3\}, \{v_2, v_3\}, \{v_3, v_5, v_6\}, \{v_4\}\}$

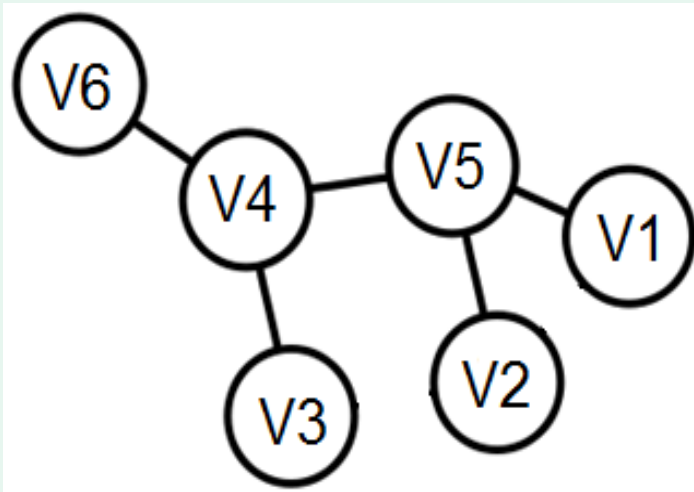
Notions préliminaires

Graphe vs Hypergraphe

l'Hypergraphe généralise la notion de Graphe. première fois par Claude Berge en 1960 [2]

Graphe

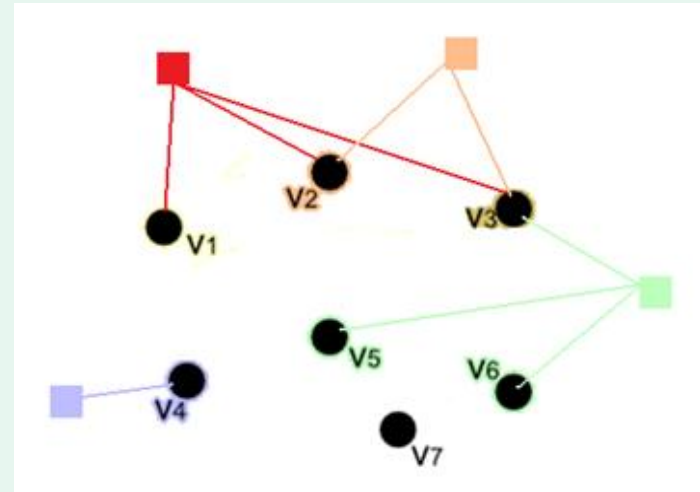
deux sommets par arête



$$\begin{aligned}
 V &= \{v_1, v_2, v_3, v_4, v_5\} \\
 E &= \{e_1, e_2, e_3, e_4, e_5\} \\
 &= \{\{v_1, v_5\}, \{v_2, v_5\}, \{v_5, v_4\}, \{v_4, v_3\}, \{v_4, v_6\}\}
 \end{aligned}$$

Hypergraphe

plusieurs sommets par hyper-arête

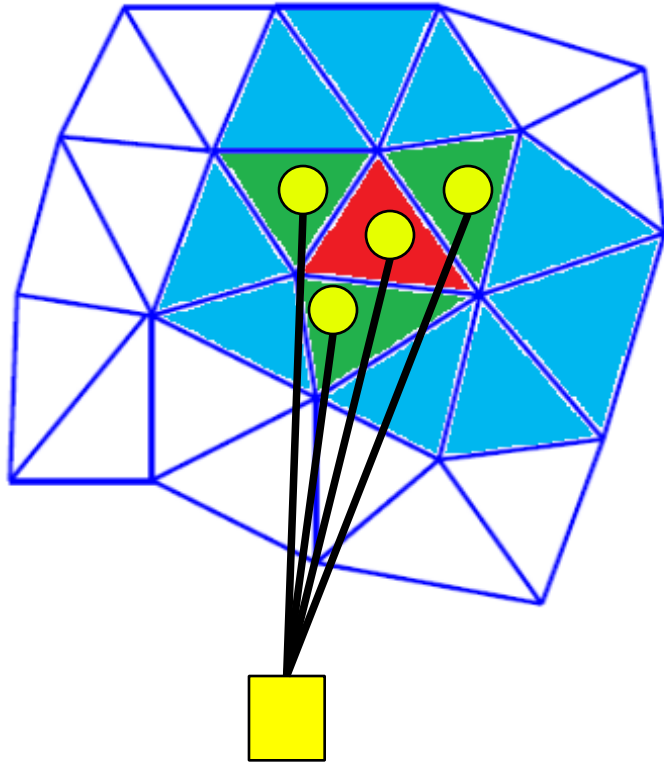


$$\begin{aligned}
 V &= \{v_1, v_2, v_3, v_4, v_5, v_6, v_7\} \\
 E &= \{h_1, h_2, h_3, h_4\} \\
 &= \{\{v_1, v_2, v_3\}, \{v_2, v_3\}, \{v_3, v_5, v_6\}, \{v_4\}\}
 \end{aligned}$$

PLAN

- Introduction
- **Modélisation**
- Résultats
- Conclusion et perspectives

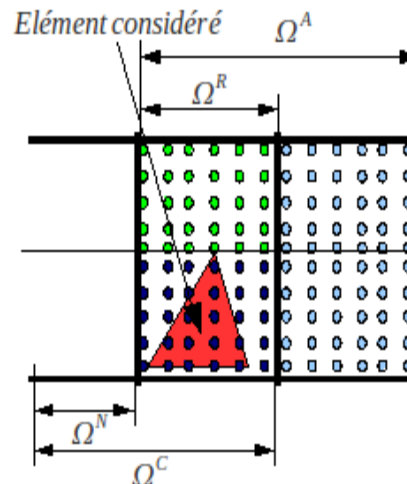
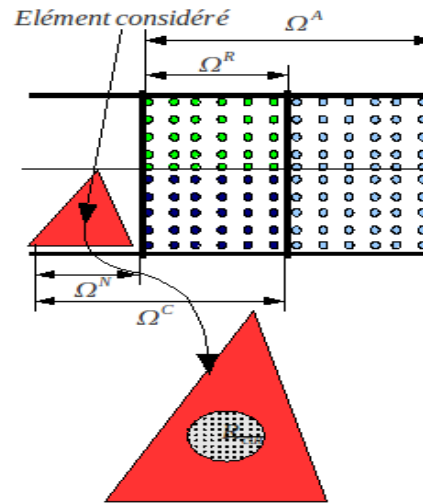
Modélisation



Coût hyper-arête de calcul

$$C_h \sim d$$

d est la dimension du problème



Coût de calcul
Sommet de la
zone normale

$$C_e = C_F$$

Coût de calcul
de la force F

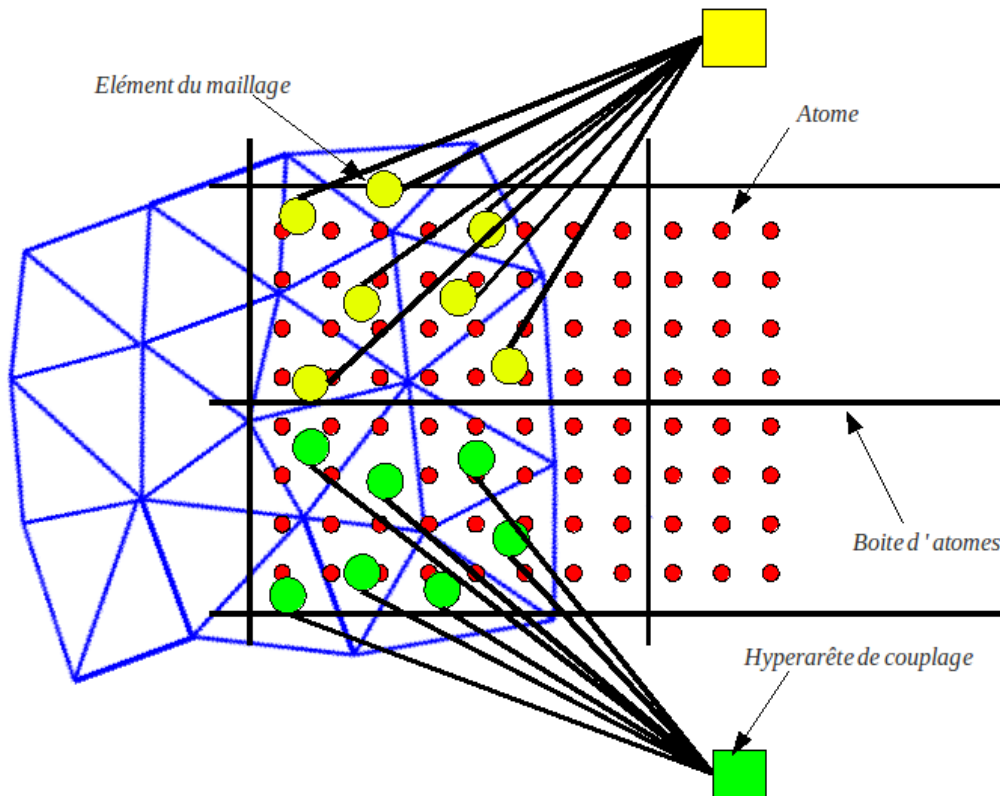
Coût de calcul
Sommet de la zone
de recouvrement

$$C_e^{BZ} = C_e + C_{coup}$$

Surcoût de couplage

Modélisation

Hyper-arête de couplage = regroupement des éléments d'une même boîte



Coût hyper-arête de couplage

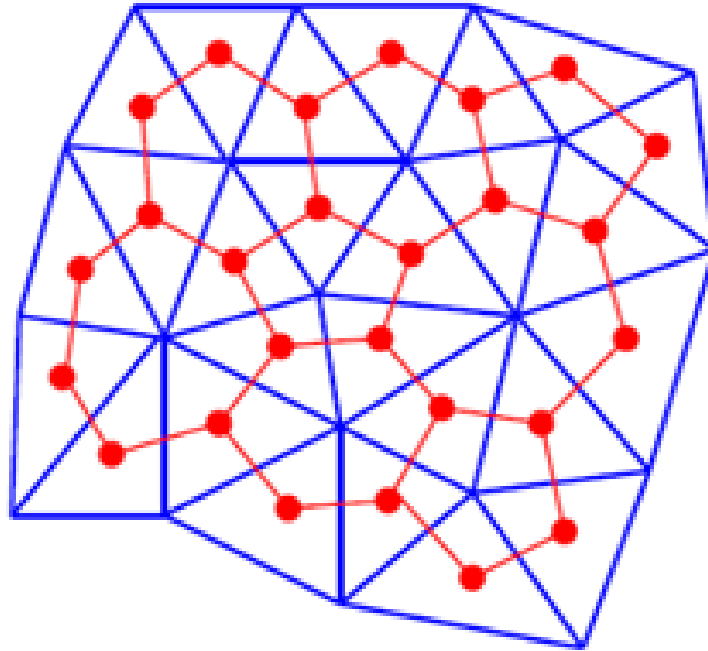
$$C_h^{Coup} \sim \gamma$$

→ Déterminé expérimentalement
d'un ordre de grandeur $> C_h$

→ Eviter de couper les éléments d'une même boîte

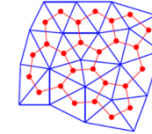
Stratégies de partitionnement

Avec modélisation en graphe

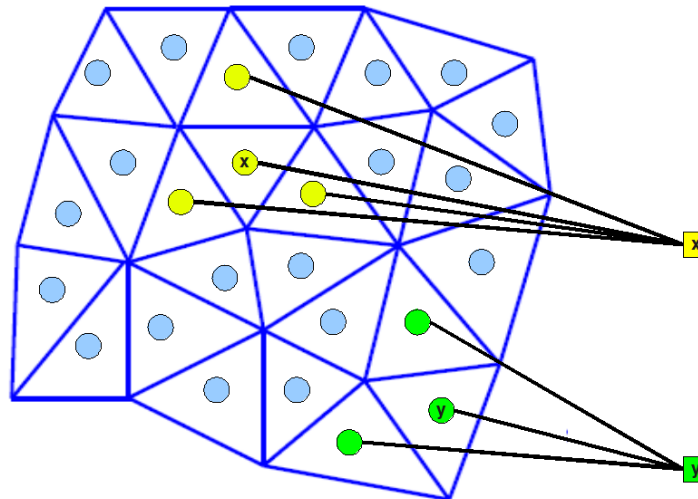


Stratégies de partitionnement

Avec modélisation en graphe

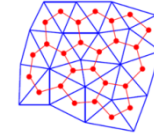


Avec modélisation en hypergraphe

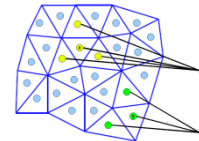


Stratégies de partitionnement

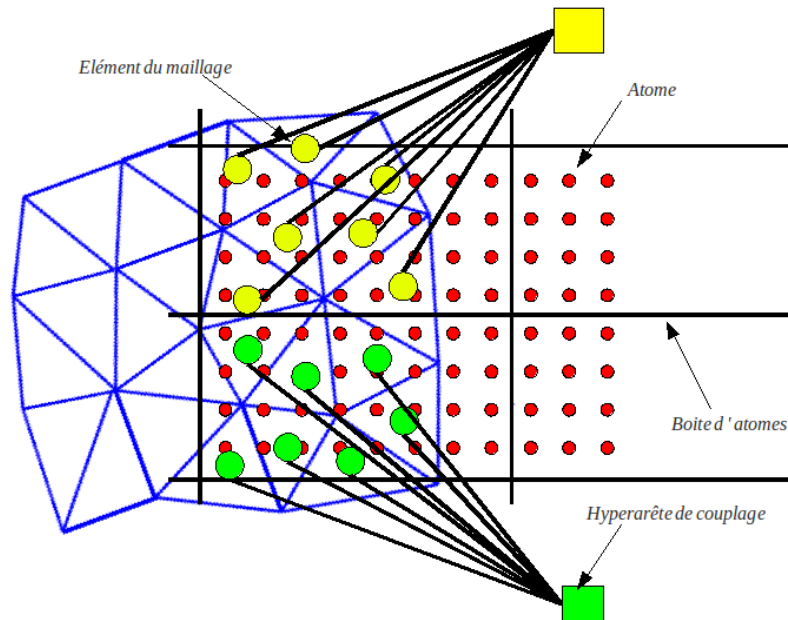
Avec modélisation en graphe



Avec modélisation en hypergraphe

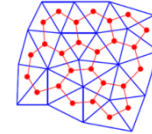


Avec modélisation en hypergraphe avec des hyper-arêtes de couplage

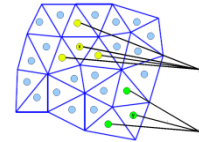


Stratégies de partitionnement

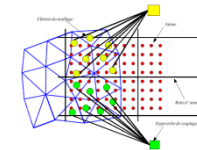
Avec modélisation en graphe



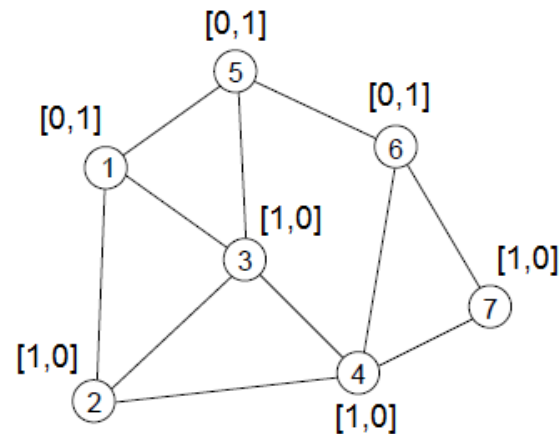
Avec modélisation en hypergraphe



Avec modélisation en hypergraphe avec des hyper-arêtes de couplage

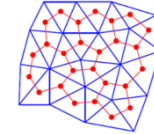


Avec modélisation en un hypergraphe avec la technique de multi-contraintes

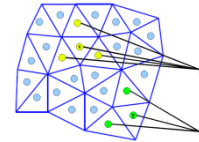


Stratégies de partitionnement

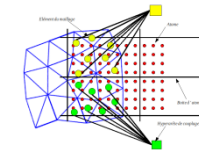
Avec modélisation en graphe



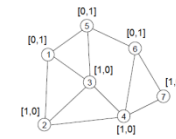
Avec modélisation en hypergraphe



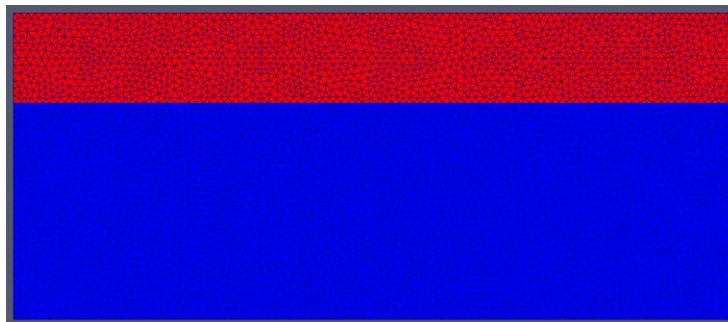
Avec modélisation en hypergraphe avec des hyper-arêtes de couplage



Avec modélisation en un hypergraphe avec la technique de multi-contraintes

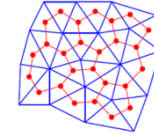


Avec modélisation en un hypergraphe avec technique des points fixes

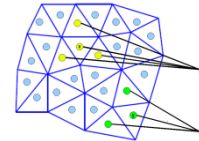


Stratégies de partitionnement

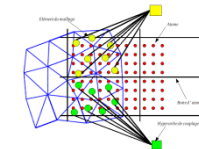
Avec modélisation en graphe



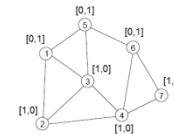
Avec modélisation en hypergraphe



Avec modélisation en hypergraphe et des hyper-arêtes de couplage



Avec modélisation en un hypergraphe avec la technique de multi-contraintes

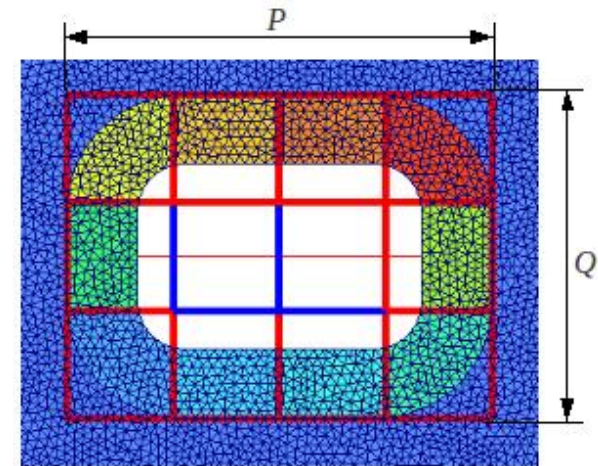
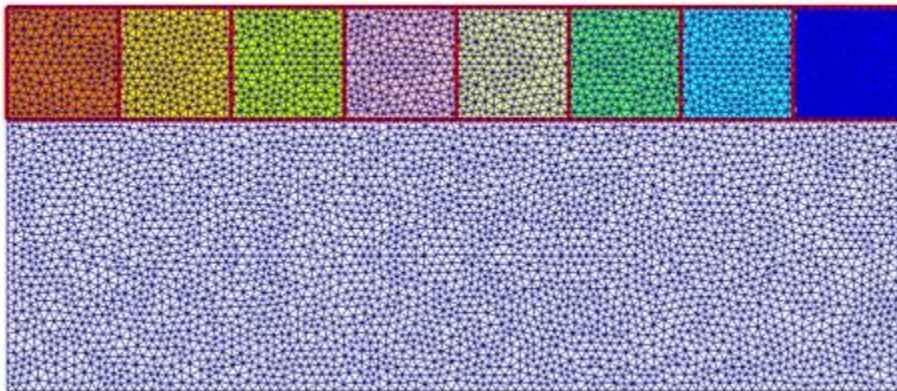


Avec modélisation en un hypergraphe avec technique des points fixes



Stratégies de partitionnement

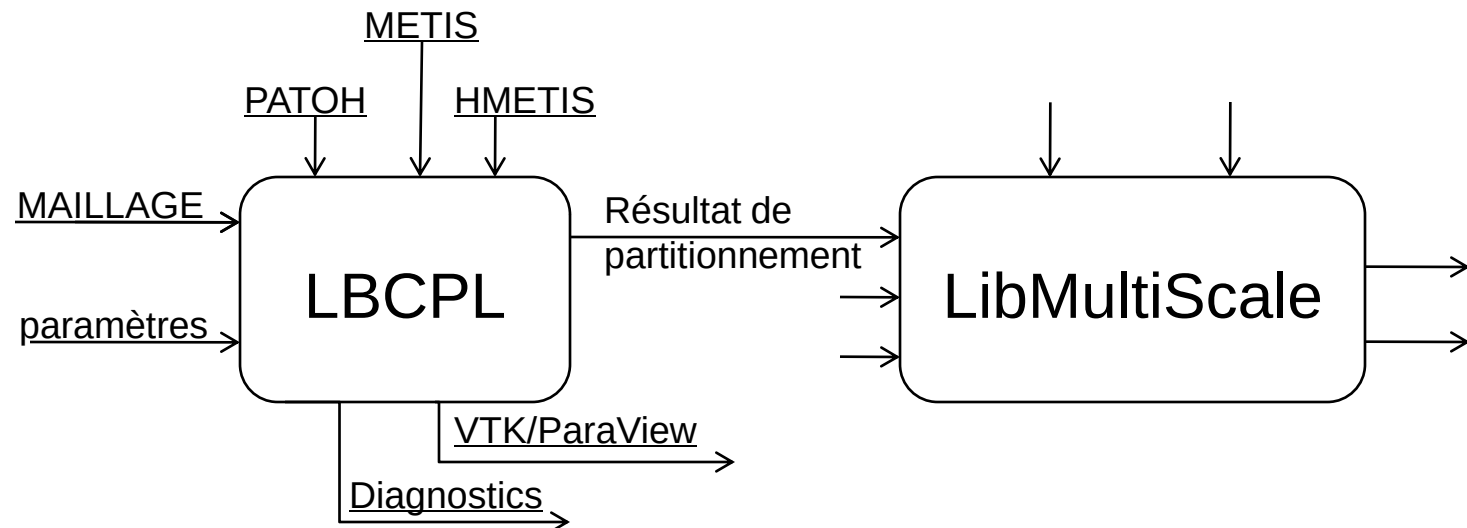
Avec modélisation en un hypergraphe avec technique des points fixes
Et hyper-arêtes de couplage



- On partitionne la zone de couplage sur les N procs,
- On partitionne le reste du maillage en gardant le premier partitionnement.

Réalisation

- **Module développé en C**
- **Utilisation de partitionneurs de graphes et d'hypergraphes**
PATOH ; METIS ; HMETIS
- **Disponible sur la gforge de l'INRIA**
<http://gforge.inria.fr/projects/mpicpl/>



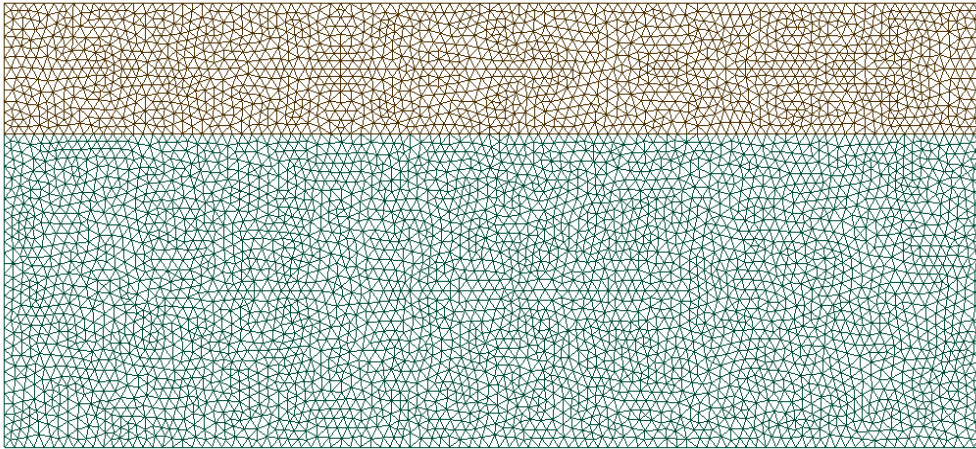
PLAN

- Introduction
- Modélisation
- **Résultats**
- Conclusion et perspectives

Résultats

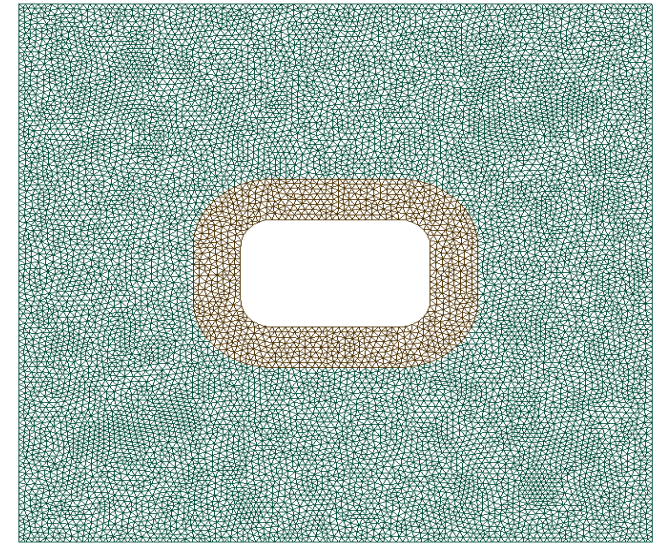
Cas test géométrie - 2D

Géométrie simple



8 procs MD $\leftrightarrow$ 8 procs FE

Géométrie complexe



12 procs MD $\leftrightarrow$ 8 procs FE

Coût sur les sommets

Calculs théoriques : surcoût de couplage négligeable devant le coût normal

→ Coût uniforme sur tous les sommets

Coût sur les hyper-arêtes de calcul

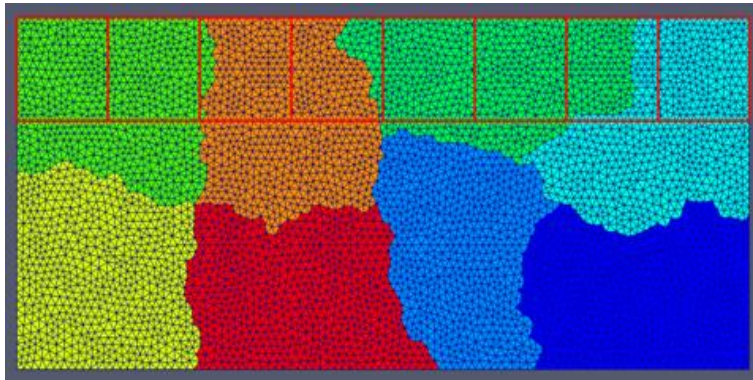
→ Constant uniforme

Coût sur les hyper-arêtes de couplage

→ Déterminé expérimentalement

Résultats

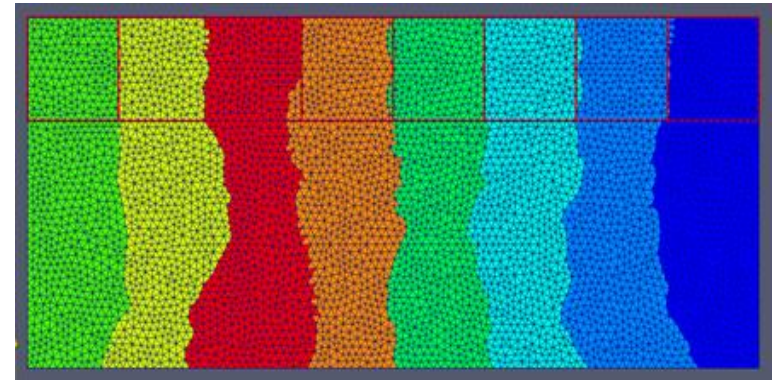
Avant



Modélisant par un **graphe**

Phase standard	12,5 % (0)
Phase de couplage	12,5 % (13,5)
com FE-MD	[0 0 2 4 3 0 2 0]

Après



Modélisant par un **hypergraphe**
points fixes et **hyper-arêtes de couplages**

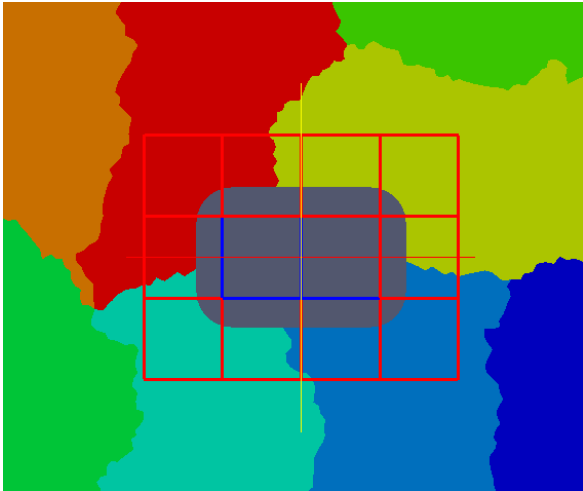
Phase standard	12.5 % (0.04)
Phase de couplage	12.5 % (0.14)
com FE-MD	[1 2 2 2 1 1 2 2]

$$C_e = 1 \quad C_e^{BZ} = 10$$

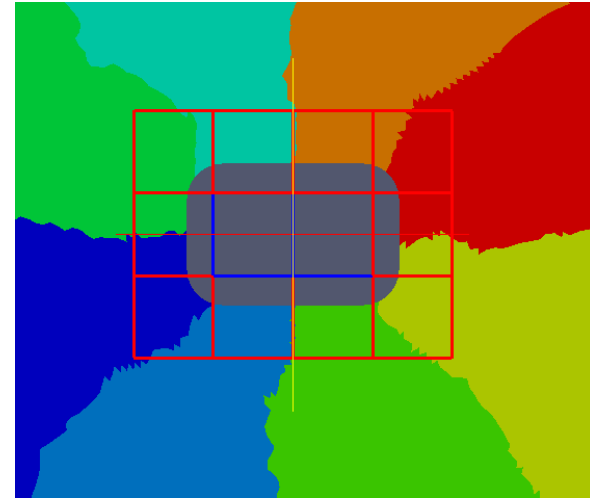
$$C_h = 1 \quad C_h^{Coup} = 50$$

Résultats

Avant



Après



Modélisant par un **graphe**

Modélisant par un **hypergraphe**
points fixes et **hyper-arêtes de couplages**

Phase standard	12,5 % (0.22)
Phase de couplage	12,5 % (12.66)
com FE-MD	[0 4 4 0 0 4 0 3]

Phase standard	12.5 % (0.08)
Phase de couplage	12.5 % (0.1)
com FE-MD	[2 3 2 2 2 2 3 2]

$$C_e = 1 \quad C_e^{BZ} = 11$$

$$C_h = 1 \quad C_h^{Coup} = 110$$

PLAN

- Introduction
- Modélisation
- Résultats
- **Conclusion et perspectives**

Conclusion et perspectives

Meilleure stratégie = **technique des points fixes + hyper-arêtes de couplage**

→ **technique des points fixes** → équilibrage entre les 2 phases → **objectif 1**

→ **hyper-arêtes de couplage** → optimisation des communications → **objectif 2**

Autres stratégies (sans points fixes)

→ Pas de différences significatives pour l'utilisation de **graphes ou d'hypergraphes**

Coûts théoriques de couplage négligeables devant les autres coûts **≠ réalité**

→ **Communications extérieures FE↔MD non prises en compte encore**

Intégration dans **LibMultiScale** *en cours*

Cas tests considérés = **géométrie en 2D** (généralement simples)

→ **structure de donnée est générique et évolutif vers des cas en 3D**

Code **source disponible en ligne** sur la gforge de l'INRIA

→ <http://gforge.inria.fr/projects/mpicpl/>

Bibliographie

- [1] G. ANCIAUX. Simulation multi-échelles des solides par une approche couplée dynamique moléculaire/éléments finis. De la modélisation à la simulation haute performance. PhD thesis, Université de Bordeaux 1 (2007)
- [2] G. ANCIAUX. LibMultiScale, AMEL User Guide. Université Bordeaux 1, septembre 2007.
- [3] C. Berge, *Introduction à la théorie des hypergraphes*, Université de Montréal (1973).
- [4] Metis - serial graph partitioning. - Technical report. glaros.dtc.umn.edu/gkhome/
- [5] Hmetis - serial hypergraph partitioning - Technical report. glaros.dtc.umn.edu/gkhome/
- [6] SANDIA. Zoltan, [http ://www.cs.sandia.gov/zoltan/](http://www.cs.sandia.gov/zoltan/). [http ://www.cs.sandia.gov/Zoltan/](http://www.cs.sandia.gov/Zoltan/).
- [7] Cevdet Aykanat Umit V. Catalyurek. PaToH : Partitioning Tool for Hypergraphs. The Ohio State University Department of Biomedical Informatics, Bilkent University Computer Engineering Department, 1999.
- [8] Umit V. Catalyurek al. Hypergraph-based dynamic load balancing for adaptive scientific computations.
- [9] Jean-François Remacle Christophe Geuzaine. Gmsh Reference Manual, 2009.
- [10] O.C. Zienkiewicz and R.L. Taylor. The nite element method, volume 1. 4 edition.
- [11] Karen D. Devine, Rob H. Bisseling, Parallel Hypergraph Partitioning for Scientific Computing, Sandia National Laboratories
- [12] Umit V. Catalyurek al. Hypergraph-based Dynamic Load Balancing for Adaptive Scientific Computations, Sandia National Laboratories (2007)
- [13] Umit V. Catalyurek , A Repartitioning Hypergraph Model for Dynamic Load Balancing, Sandia National Laboratories (2008)
- [14] Karen D. Devine, al. Balancing for Emerging Parallel Applications and Architectures, Sandia National Laboratories
- [15] James D. Teresco , al. Partitioning and Dynamic Load Balancing for the Numerical Solution of Partial Differential Equations, Sandia National Laboratories

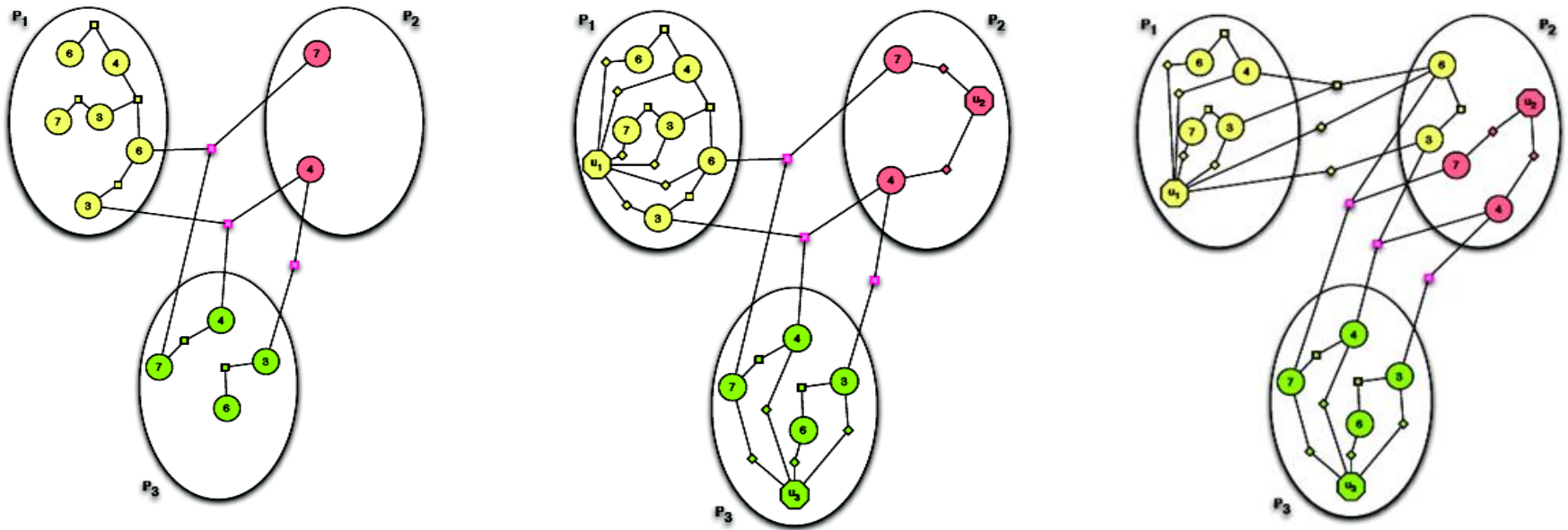
Questions ?

ANNEXE A état de l'art

Comment il ont fait ? [11 ; 12 ; 13 ; 14 ; 15]	Qu'est ce qu'on veut faire ? Et comment adapter ?
• Modélisation par hypergraphe → Estimation précise des coûts de calcul	• Estimation précise des coûts → Modélisation par hypergraphe
• Équilibrage dynamique de la charge dans le même code, • Répartition entre plusieurs processeurs • Repartitionnement dynamique de l'hypergraphe	• Équilibrage statique de charge entre deux codes • Répartition entre plusieurs processeurs • Un seul partitionnement de l'hypergraphe à l'initiation
• coût de migration des données	• coût de couplage

ANNEXE A état de l'art

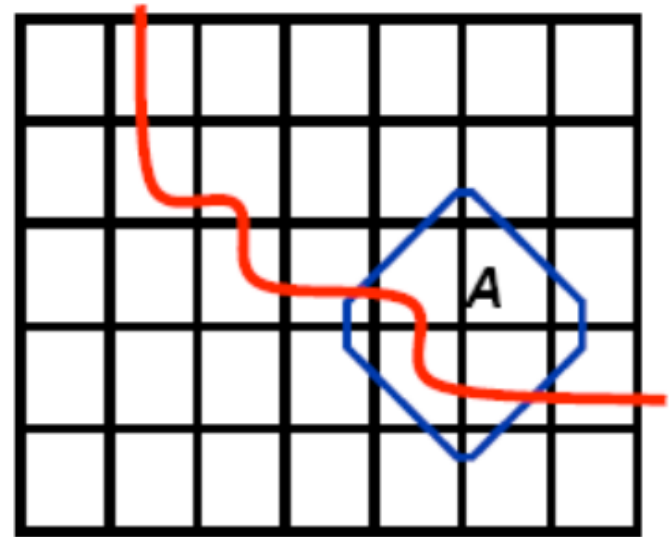
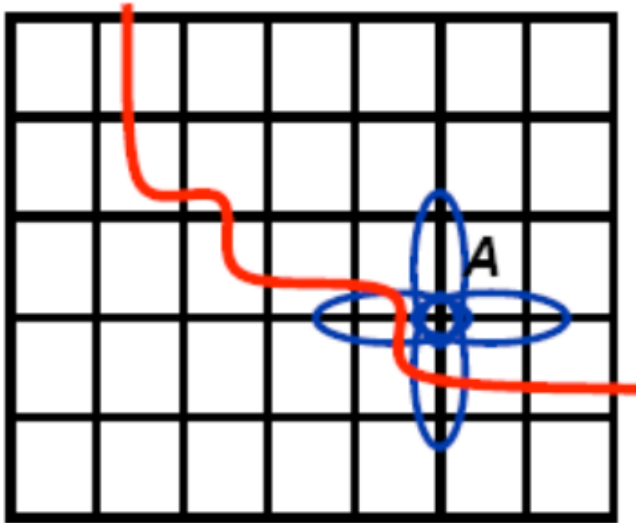
Équilibrage de charge dynamique



[13] Umit V. Catalyurek al.

ANNEXE A *Avantage de l'hypergraphe*

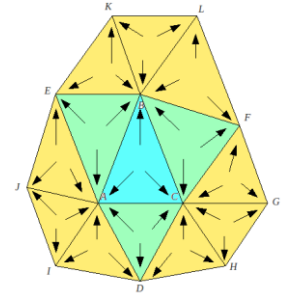
➔ Regrouper plusieurs sommets ayant une propriété commune ensemble



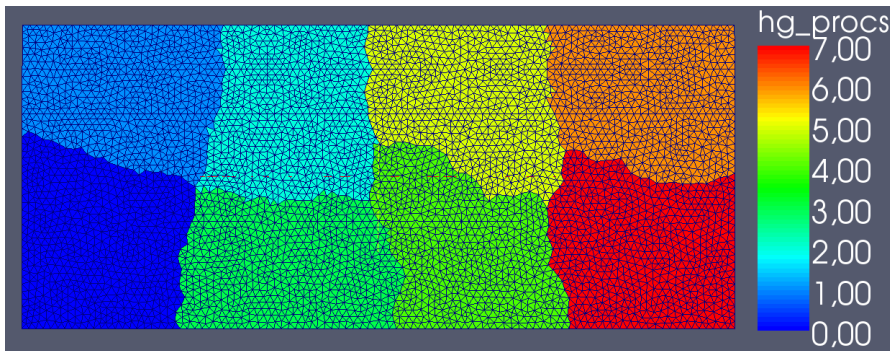
➔ Modélisation réelle des coûts de communications

[14] <http://www.cs.sandia.gov/Zoltan>

ANNEXE C



simplification = considérer uniquement les premiers voisins



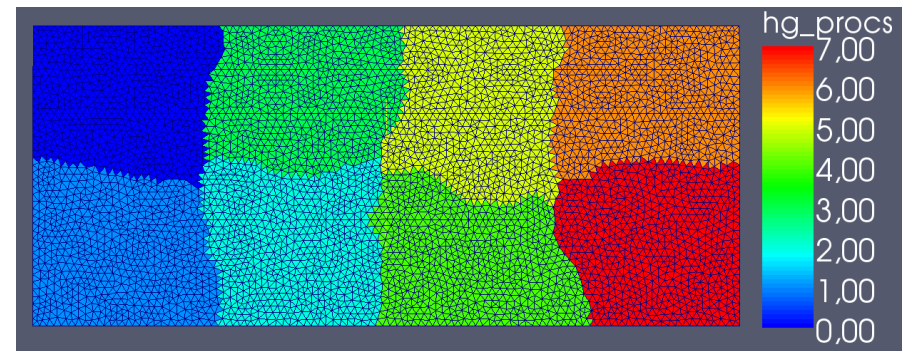
en considérant les premiers voisins.

nb hyperedges = 9542

total weights = 9542

edgecut = 445

% weights = [12 13 13 13 13 12 13 12]



En considérant les premiers et seconds voisins.

nb hyperedges = 9542

total weights = 9542

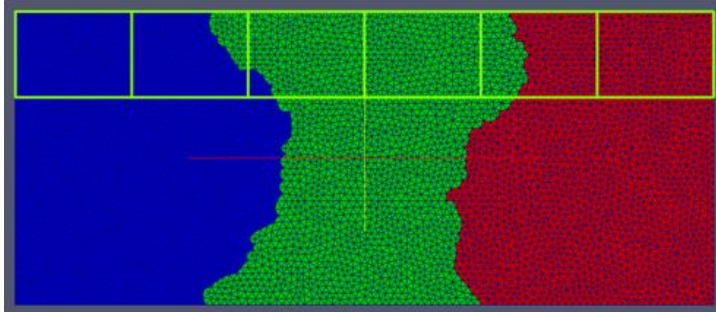
edgecut = 959

% weights = [13 13 13 13 12 12 12 12]

➔ dans le second cas, les hyper-arêtes se chevauchent davantage, mais ça n'a pas d'influence à priori pour le partitionneur

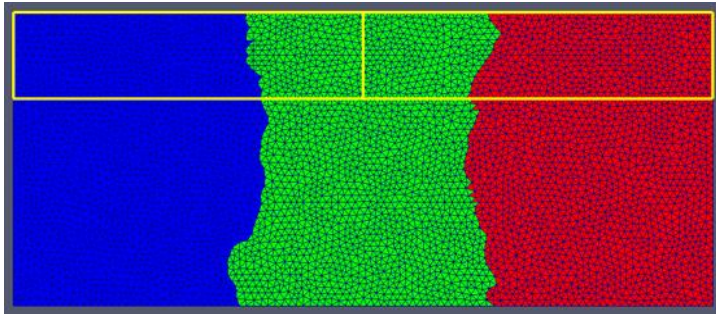
ANNEXE D

Avantages du modèle



Avant

Comm FE-MD : [3 4 2]
% répart ph1 : [34 33 33]
% répart ph2 : [32 40 28]

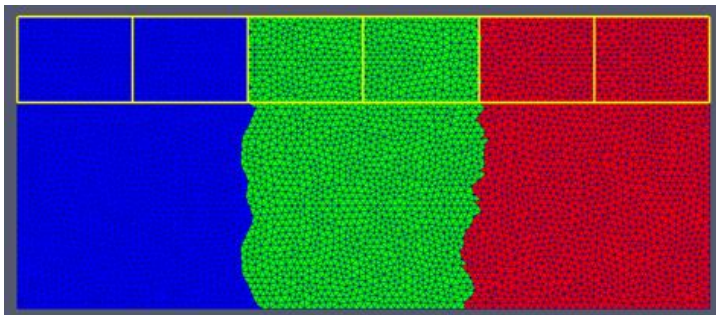
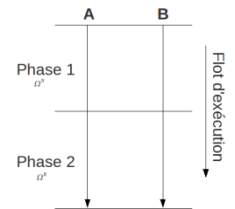


Après

Equilibrer la charge pendant les deux phases

→ partitionnement avec points fixes

% répart ph1 : [33 33 33]
% répart ph2 : [33 33 33]



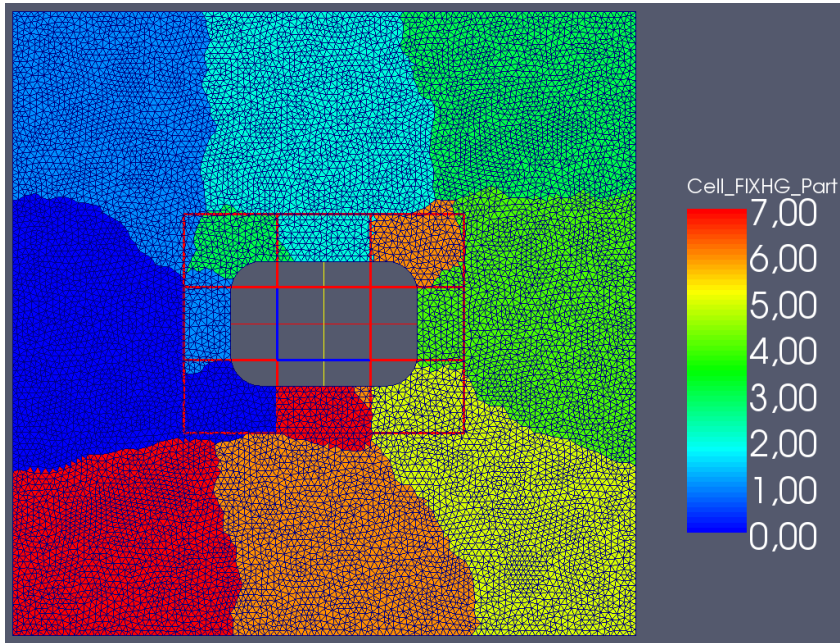
Améliorer les communications FE-MD

→ partitionnement avec hyper-arêtes de couplage

Comm FE-MD : [2 2 2]
% répart ph1 : [33 33 33]
% répart ph2 : [33 33 33]

ANNEXE D

Limitations du modèle



nous ne pouvons pas assurer que les parties accordées à un même processeur soient contiguës sur le maillage.

Avec un coût uniforme pour les sommets et les hyper-arêtes et avec des hyper-arêtes de couplages.

coût He

cHe	N	B	N	B
50	1	1	1	1

poids v

% répart ph1 : [12 12 13 13 12 13 13 13]

% répart ph2 : [13 13 13 12 13 12 12 13]

com FE-FE : [3 3 3 4 3 3 6 3]

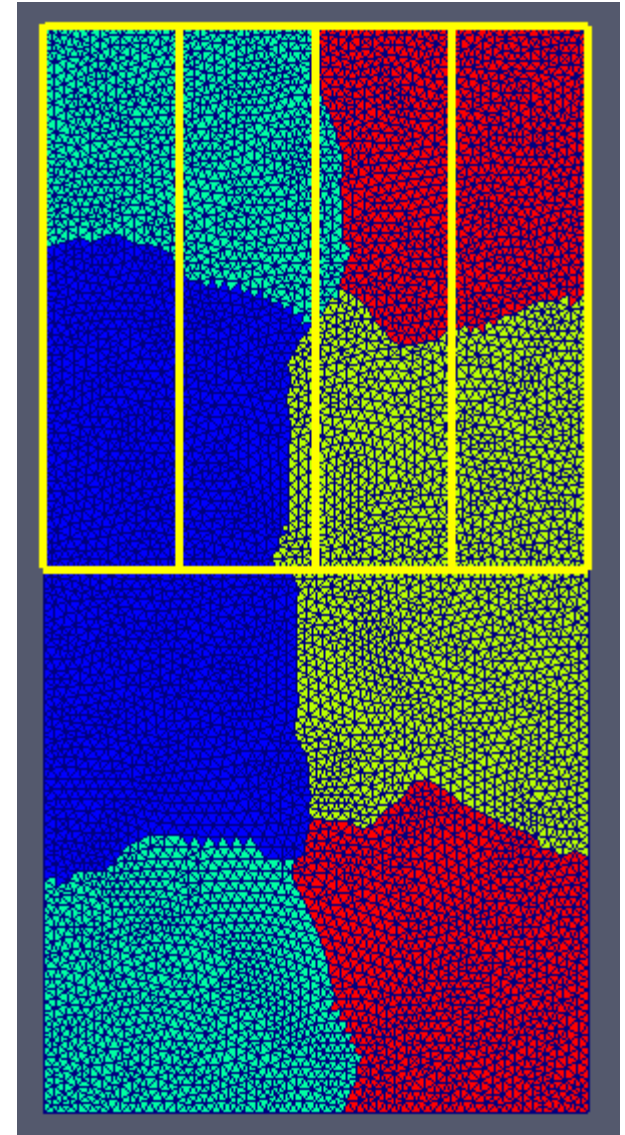
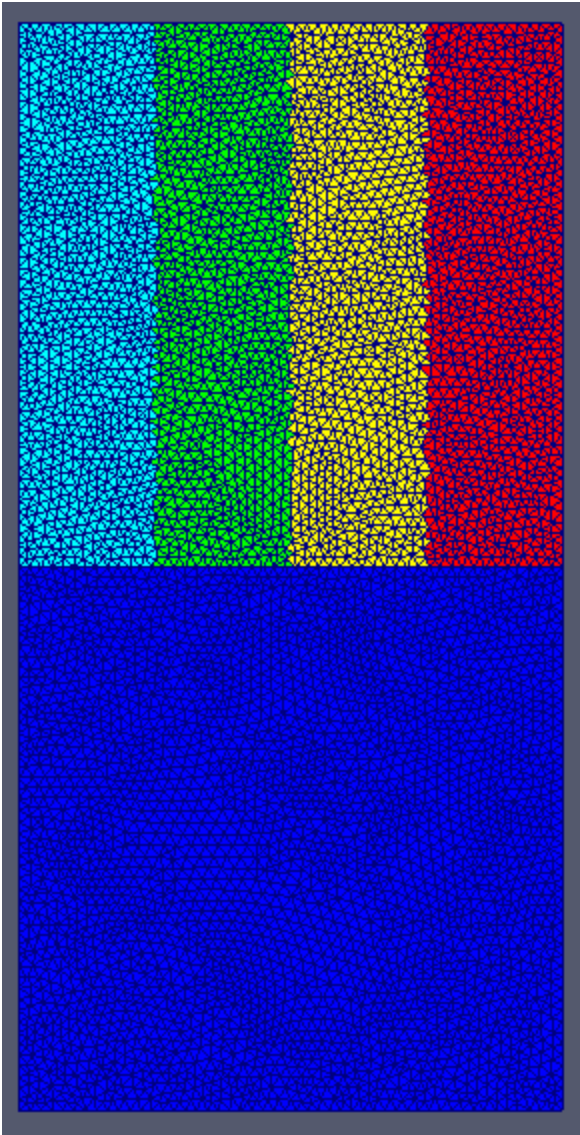
com FE-MD : [1 3 2 2 3 1 2 1]

Equilibrage du poids

ph1	ph2
12.5 ; 0.1	12.5 ; 0.12

ANNEXE D

Limitations du modèle



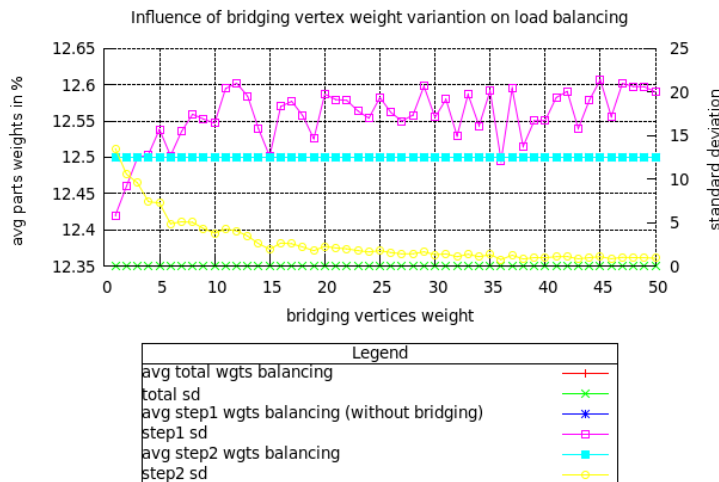
ANNEXE E détermination expérimentale du coût

Calculs des coûts en fonction des complexités et des dépendances,
Mais on a mis des coût artificiels pour mieux guider le partitionneur.

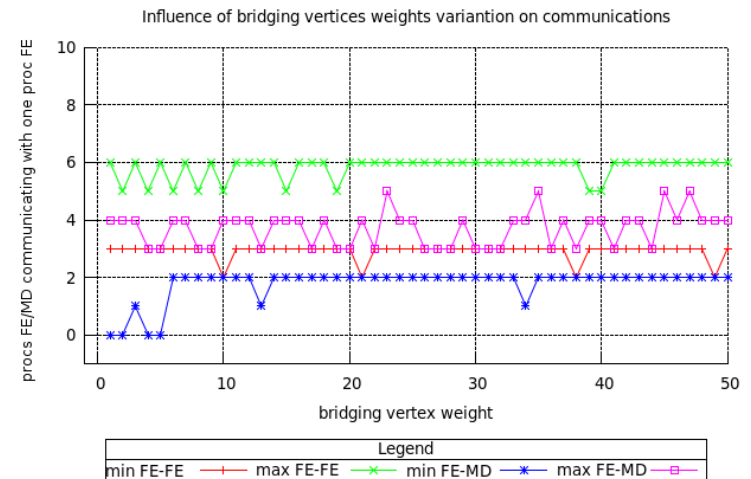
Comment les déterminer ?

➔ une étude exhaustive en faisant varier les coûts et voir leur influence
sur les résultats de partitionnement

Cas d'une modélisation avec un graphe



Influence de la variation du poids des sommets de la zone de recouvrement du graphe sur l'équilibrage de charge.

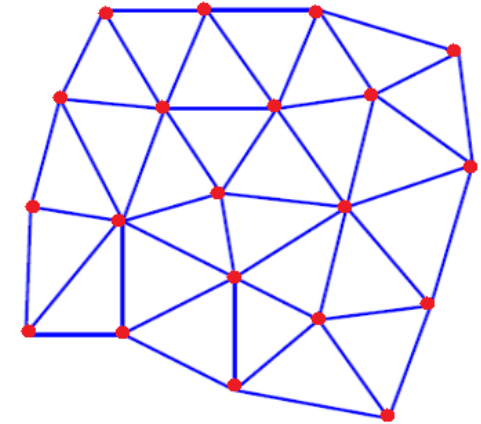
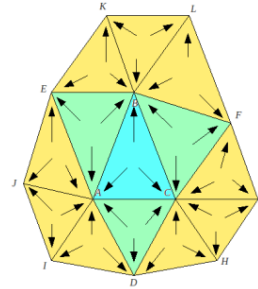


Influence de la variation du poids des sommets de la zone de recouvrement du graphe sur les communications.

ANNEXE F perspective désynchronisation

tâche d'intégration FE

$$(T_S^C) \left\{ \begin{array}{l} \dot{u}_I^{n+1/2} = \dot{u}_I^n + \frac{\Delta t}{2M_I} F_I^n \\ u_I^{n+1} = u_I^n + \Delta t \dot{u}_I^{n+1/2} \\ \text{Calcul de } F_I^{n+1} \\ \dot{u}_I^{n+1} = \dot{u}_I^{n+1/2} + \frac{\Delta t}{2M_I} F_I^{n+1} \end{array} \right.$$



- Synchronisation nécessaire à chaque pas de temps.
- Possibilité de fusionner la dernière étape avec l'étape de calcul de la force F

Algorithme 1 Algorithme de calcul de la force F

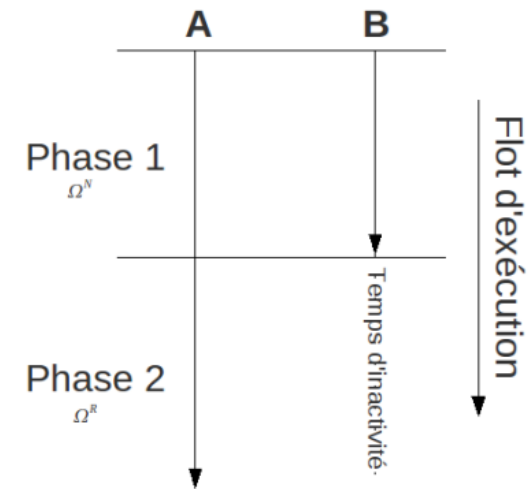
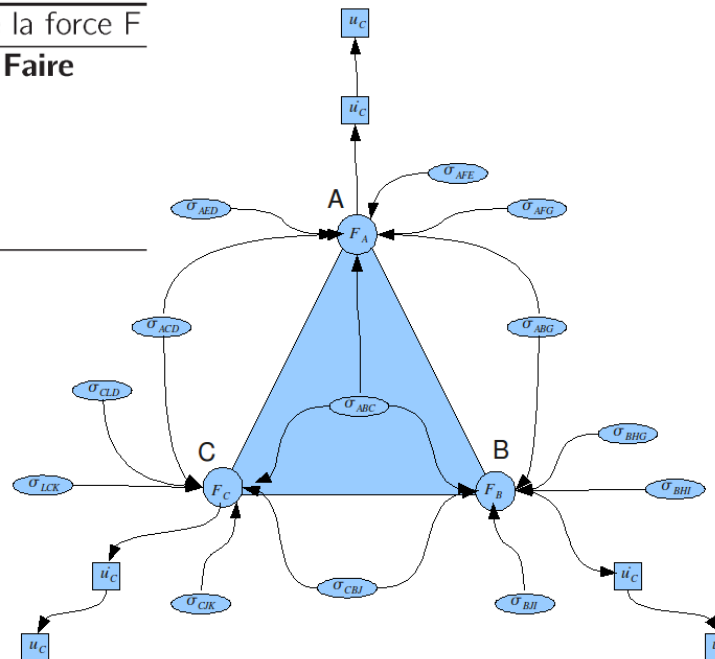
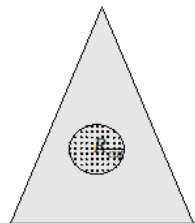
Pour tout les éléments du maillage Faire

déformation-crystal()

calcul-potentiel()

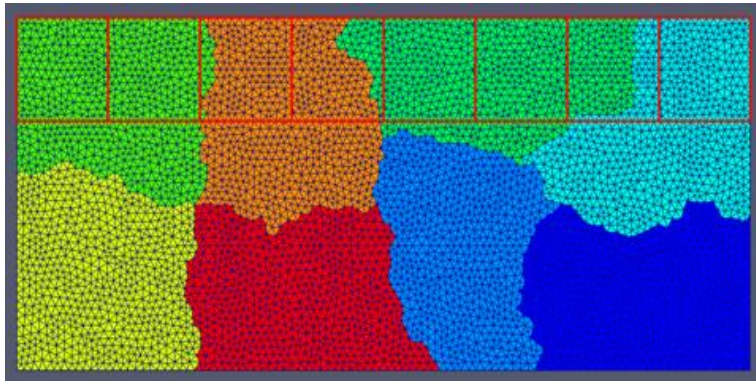
pondération-sommets()

FinPour



ANNEXE G autres résultats

Partitionnement du maillage en le modélisant par un **graphe**



Coût uniforme

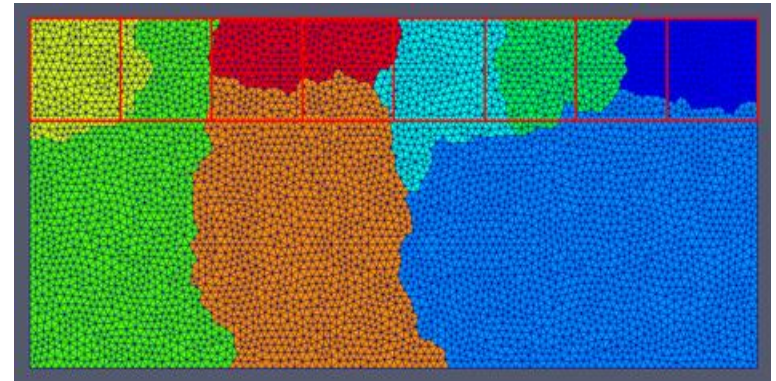
1 1 1 1 12,5 ; 0 12,5 ; 13,5

% répart ph1 : [12 13 13 13 12 13 13 13]

% répart ph2 : [0 0 17 36 26 0 21 0]

com FE-FE : [3 6 4 4 3 3 5 4]

com FE-MD : [0 0 2 4 3 0 2 0]



Surcoût pour les sommets

1 1 1 10 12.5 ; 0,01 12,5 ; 3,8

% répart ph1 : [13 12 12 12 12 13 12 13]

% répart ph2 : [15 5 15 15 10 15 9 15]

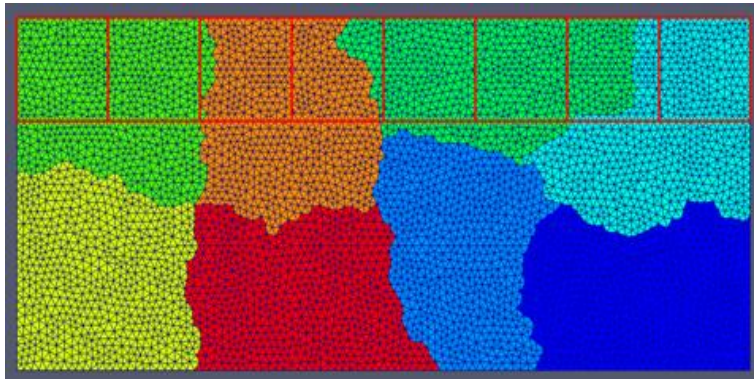
com FE-FE : [3 5 5 4 4 2 5 4]

com FE-MD : [2 3 3 2 2 2 3 4]

ANNEXE G autres résultats

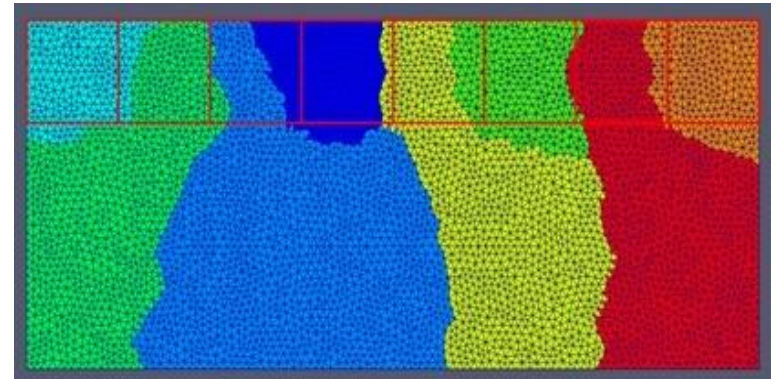
Partitionnement du maillage en le modélisant par un **hypergraphe**

Avant



Coût uniforme

Après



Surcoût pour les sommets et les hyper-arêtes

1 1 1 1 12,5 ; 0 12,5 ; 13,5

0 1 10 1 10 12.5 ; 0.02 12.5 ; 2.8

% répart ph1 : [12 13 13 13 12 13 13 13]

% répart ph2 : [0 0 17 36 26 0 21 0]

com FE-FE : [3 6 4 4 3 3 5 4]

com FE-MD : [0 0 2 4 3 0 2 0]

% répart ph1 : [12 12 12 12 13 13 13 13]

% répart ph2 : [15 7 15 11 15 10 15 11]

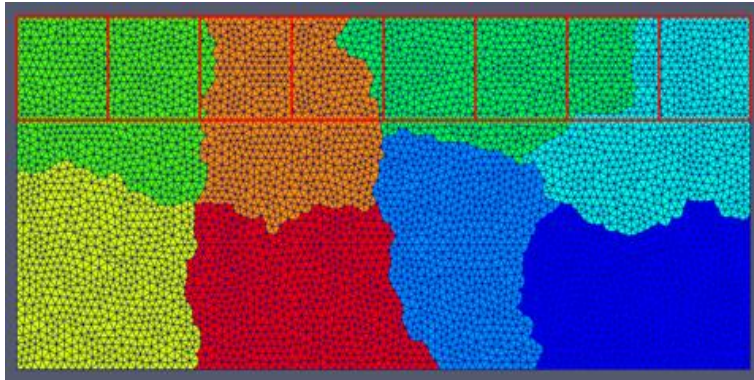
com FE-FE : [2 3 1 2 2 4 1 3]

com FE-MD : [2 2 2 2 3 2 2 2]

ANNEXE G autres résultats

Partitionnement du maillage en le modélisant par un **hypergraphe** avec des **hyper-arêtes de couplage**

Avant



Coût uniforme

1 1 1 1 12,5 ; 0 12,5 ; 13,5

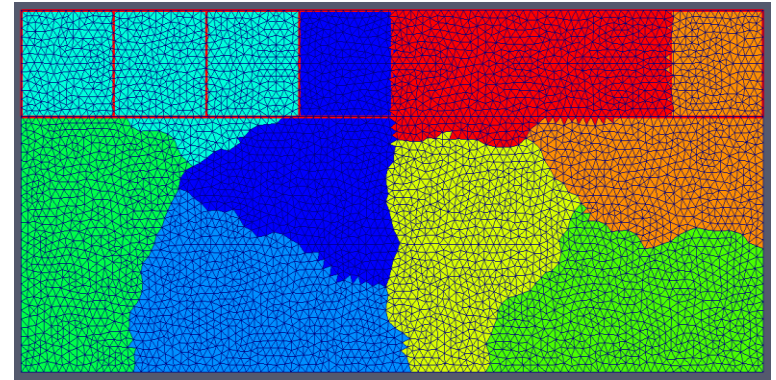
% répart ph1 : [12 13 13 13 12 13 13 13]

% répart ph2 : [0 0 17 36 26 0 21 0]

com FE-FE : [3 6 4 4 3 3 5 4]

com FE-MD : [0 0 2 4 3 0 2 0]

Après



Coût uniforme

100 1 1 1 1 12.5 ; 0.13 12.5 ; 15.26

% répart ph1 : [12 13 13 12 12 13 12 13]

% répart ph2 : [12 0 38 0 0 0 13 37]

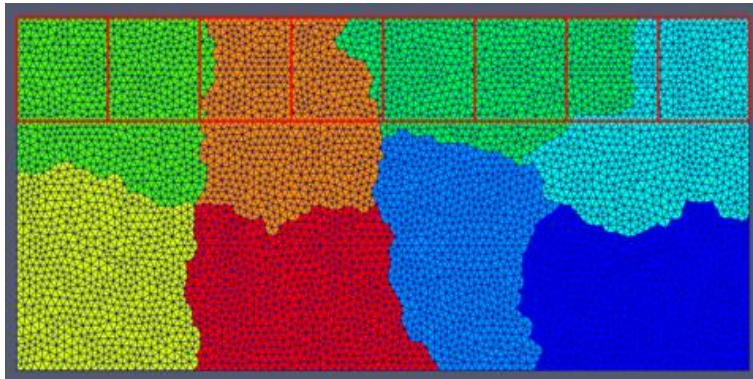
com FE-FE : [5 3 2 3 2 5 3 3]

com FE-MD : [1 0 3 0 0 0 1 3]

ANNEXE G autres résultats

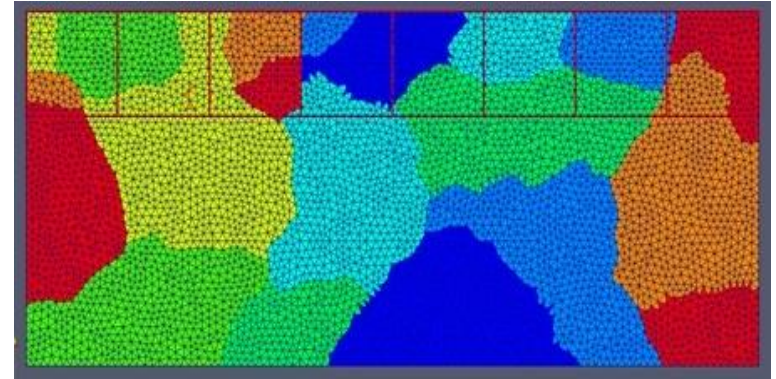
Partitionnement du maillage en le modélisant par un **hypergraphe** avec la technique de **multi-contraintes**

Avant



Coût uniforme

Après



Technique multi-contraintes

1 1 1 1 12,5 ; 0 12,5 ; 13,5

12.5 ; 0.22 12.5 ; 0.27

% répart ph1 : [12 13 13 13 12 13 13 13]

% répart ph2 : [0 0 17 36 26 0 21 0]

com FE-FE : [3 6 4 4 3 3 5 4]

com FE-MD : [0 0 2 4 3 0 2 0]

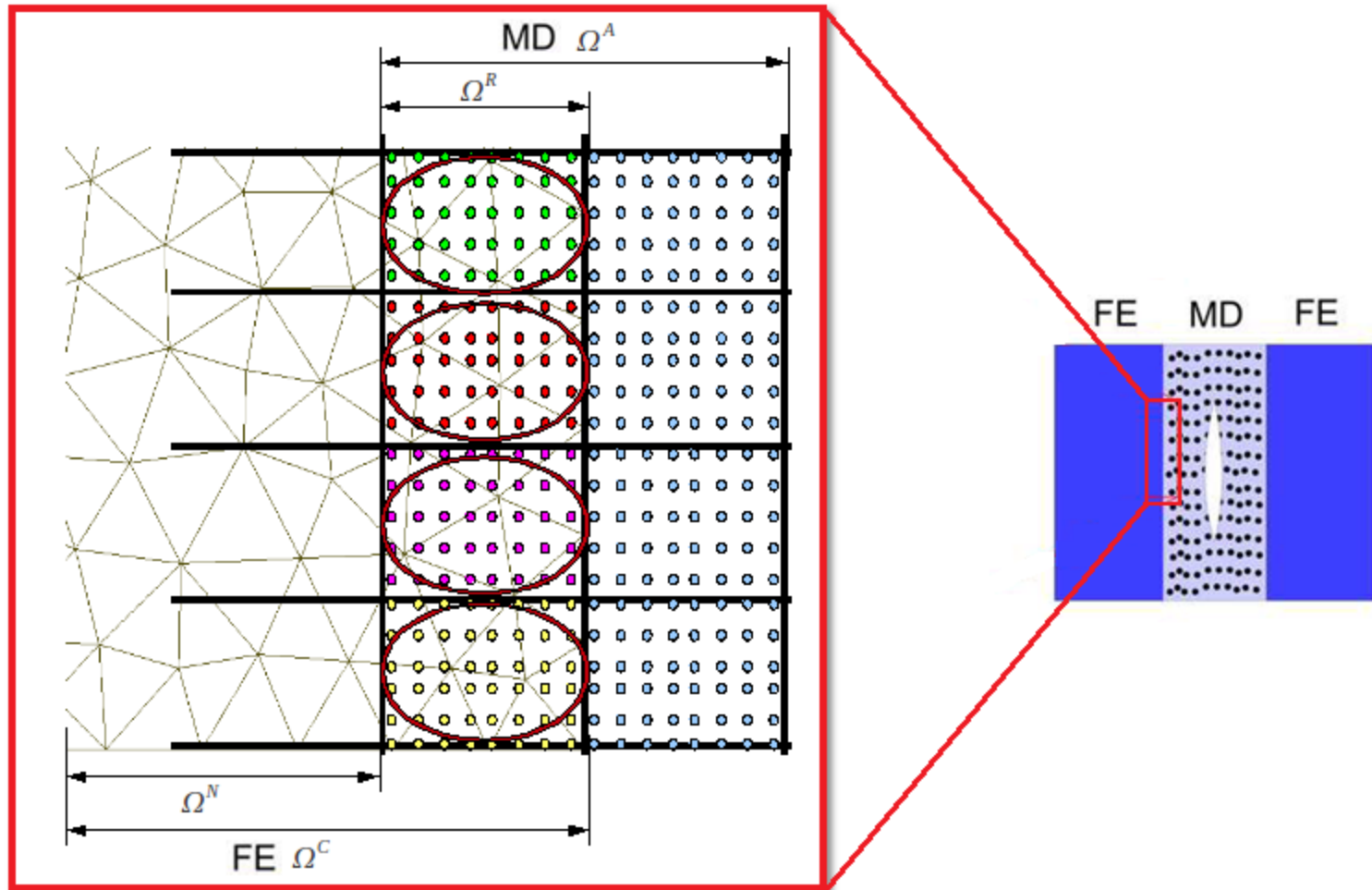
% répart ph1 : [12 12 12 12 13 12 13 13]

% répart ph2 : [12 12 12 13 12 12 13 13]

com FE-FE : [4 5 6 5 5 4 5 6]

com FE-MD : [2 4 4 3 2 3 6 3]

ANNEXE H modélisation



ANNEXE I Réalisation

Environnement d'exécution de notre module

