

# Adéquation algorithme architecture pour l'accélération de méthodes d'inversion de données en grande dimension

Nicolas GAC

Maître de conférences à l'université Paris Saclay  
Laboratoire des Signaux et Systèmes (L2S) - Groupe Problèmes Inverses (GPI)



*Soutenance HDR, 23 novembre 2020*

- 1 Parcours depuis le doctorat
- 2 Adéquation Algorithme Architecture ( $A^3$ )
- 3 Calcul intensif sur GPU/FPGA
- 4 Calcul distribué sur serveur multiGPU
- 5 Perspectives

## Formation

2000-04 ENSIMAG + DEA Microélectronique + M2 Trait. signal

2004-08 **Thèse** (GIPSA-lab)

*A<sup>3</sup> sur FPGA pour la tomographie TEP*

[R<sub>2009</sub>, R<sub>2012</sub>]

2007-08 ATER à l'Université de Cergy

2008-09 **Post-Doctorat** (L2S/CEA-List)

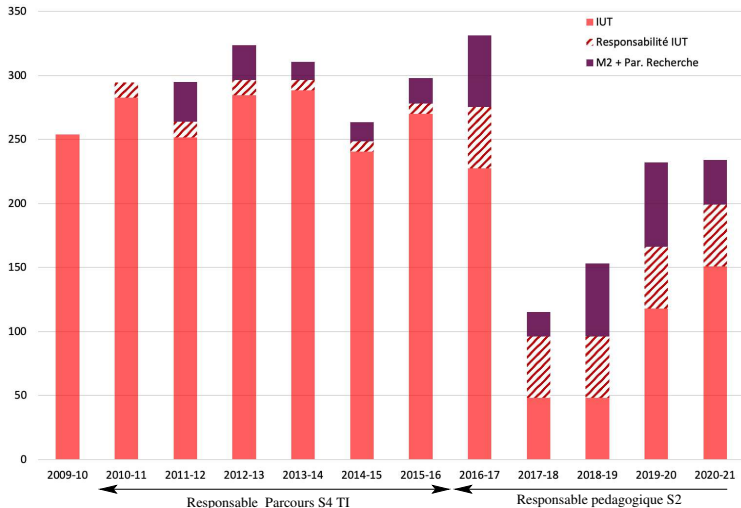
*A<sup>3</sup> sur GPU pour la tomographie X*

2009-.. **Maître de conférences** à l'Université Paris Saclay (IUT de Cachan)

**Recherche** au Groupe Problèmes Inverses (GPI)

2017-19 **Délégation CNRS** à mi-temps au laboratoire Lagrange (Nice)

*Parallélisation multiGPU pour la radioastronomie*



Enseignements IUT (90%) et M2 (10 %)

**IUT** Systèmes numériques embarqués

**M2** M2 SETI (parallélisation GPU/FPGA) + M2 ATSI (Imagerie médicale)



- 1 Parcours depuis le doctorat
- 2 Adéquation Algorithme Architecture ( $A^3$ )
  - Problématique et positionnement
  - $A^3$  au sein du GPI
  - Encadrement de thèses et collaborations
- 3 Calcul intensif sur GPU/FPGA
  - Paire de projection/rétroprojection  $\mathcal{P}/\mathcal{R}$
  - Opérateur de convolution  $\mathcal{C}$
- 4 Calcul distribué sur serveur multiGPU
  - Reconstruction tomographique
  - DeconvSKA : Déconvolution en radioastronomie
- 5 Perspectives
  - Une méthodologie  $A^3$  à échelle humaine
  - SKA, un défi  $A^3$  à grande échelle

# Adéquation Algorithme Architecture [*thème C du GDR ISIS*]

## Meilleure **combinaison d'architectures et d'algorithmes**

sous **contraintes** de :

- temps de traitement
- précision de calcul
- consommation énergétique

# Adéquation Algorithme Architecture [*thème C du GDR ISIS*]

## Meilleure **combinaison d'architectures et d'algorithmes**

sous **contraintes** de :

- temps de traitement
- précision de calcul
- consommation énergétique
- coût du système de calcul
- temps de développement

# Adéquation Algorithme Architecture [thème C du GDR ISIS]

## Meilleure combinaison d'architectures et d'algorithmes

sous **contraintes** de :

- temps de traitement
- précision de calcul
- consommation énergétique
- coût du système de calcul
- temps de développement

## Différentes approches A<sup>3</sup>

- outils pour le TdSI (ex : Syndex, PREESM...)
- proche des capteurs (*smart sensors*)
- proche de l'application et des architectures

# Adéquation Algorithme Architecture [thème C du GDR ISIS]

## Meilleure combinaison d'architectures et d'algorithmes

sous **contraintes** de :

- temps de traitement
- précision de calcul
- consommation énergétique
- coût du système de calcul
- temps de développement

## Différentes approches A<sup>3</sup>

- outils pour le TdSI (ex : Syndex, PREESM...)
- proche des capteurs (*smart sensors*)
- **proche de l'application et des architectures**
  - contextes applicatifs **spécifiques avec contraintes fortes**
  - **solutions « sur mesure »** pour des performances optimales.

# Adéquation Algorithme Architecture

## Enjeux

- **Interaction humaine** la plus forte possible entre expertises :
  - **applicative**
  - **algorithmique**
  - **matérielle**
- Prise en compte des **évolutions technologiques**
- Difficile mais impérative **généralisation de la démarche A<sup>3</sup>**

# Périmètre applicatif

## Problème inverse

Reconstruction à partir des mesures d'un instrument

## Verrous

- #1 Problèmes dit « mal posé »
- #2 Problème en grande dimension

# Périmètre applicatif

## Problème inverse

Reconstruction à partir des mesures d'un instrument

## Verrous

- #1 Problèmes dit « mal posé »
- #2 Problème en grande dimension

## Méthodes analytiques

Modélisation souvent réductrice

😊 rapide ☹️ moins robuste



# Périmètre applicatif

## Problème inverse

Reconstruction à partir des mesures d'un instrument

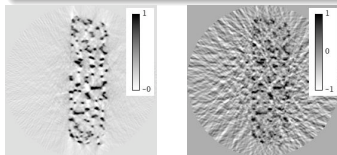
## Verrous

- #1 Problèmes dit « mal posé »
- #2 Problème en grande dimension

## Méthodes analytiques

Modélisation souvent réductrice

😊 rapide ☹️ moins robuste



FDK  
(256/256 proj.)

FDK  
(32/256 proj.)

*Reconstructions d'une  
mousse métallique [C<sub>2011a</sub>]*

# Périmètre applicatif

## Problème inverse

Reconstruction à partir des mesures d'un instrument

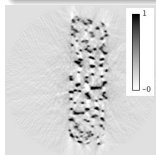
## Verrous

- #1 Problèmes dit « mal posé »
- #2 Problème en grande dimension

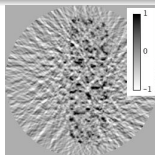
## Méthodes analytiques

Modélisation souvent réductrice

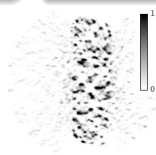
😊 rapide ☹️ moins robuste



FDK  
(256/256 proj.)



FDK  
(32/256 proj.)



Itératif  
(32/256 proj.)

## Méthodes itératives par optimisation d'un critère

Modélisation plus riche

😊 plus robuste ☹️ lent

*Reconstructions d'une  
mousse métallique [C<sub>2011a</sub>]*

# Larges champs d'exploration algorithmique

$$\mathbb{M}_D : \mathbf{g} = \mathbf{H}\mathbf{f} + \epsilon$$

$\mathbf{g}$  : mesures de l'instrument

$\mathbf{H}$  : modèle d'acquisition

$\mathbf{f}$  : objet à reconstruire

$\epsilon$  : incertitudes sur  $\mathbf{g}$  et  $\mathbf{H}$

# Larges champs d'exploration algorithmique

$$\mathbb{M}_D : \mathbf{g} = \mathbf{H}\mathbf{f} + \epsilon$$

$\mathbf{g}$  : mesures de l'instrument

$\mathbf{H}$  : modèle d'acquisition

$\mathbf{f}$  : objet à reconstruire

$\epsilon$  : incertitudes sur  $\mathbf{g}$  et  $\mathbf{H}$

## Approches bayésiennes

$$\text{Loi a posteriori : } p(\mathbf{f}|\mathbf{g}) = \frac{p(\mathbf{g}|\mathbf{f}; \mathbb{M}_D)p(\mathbf{f}; \mathbb{M}_f)}{p(\mathbf{g})}$$

- 1 Choix de l'estimateur
- 2 Choix du modèle de l'instrument  $\mathbf{H}$
- 3 Choix du modèle des incertitudes  $\epsilon$
- 4 Choix du modèle de l'objet  $\mathbb{M}_f$
- 5 Choix de l'algorithme d'optimisation

# Larges champs d'exploration algorithmique

$$\mathbb{M}_D : \mathbf{g} = \mathbf{H}\mathbf{f} + \epsilon$$

$\mathbf{g}$  : mesures de l'instrument

$\mathbf{H}$  : modèle d'acquisition

$\mathbf{f}$  : objet à reconstruire

$\epsilon$  : incertitudes sur  $\mathbf{g}$  et  $\mathbf{H}$

## Approches bayésiennes

$$\text{Loi a posteriori : } p(\mathbf{f}|\mathbf{g}) = \frac{p(\mathbf{g}|\mathbf{f}; \mathbb{M}_D)p(\mathbf{f}; \mathbb{M}_f)}{p(\mathbf{g})}$$

- 1 **Choix de l'estimateur**
  - MAP
  - EAP
- 2 **Choix du modèle de l'instrument  $\mathbf{H}$**
- 3 **Choix du modèle des incertitudes  $\epsilon$**
- 4 **Choix du modèle de l'objet  $\mathbb{M}_f$**
- 5 **Choix de l'algorithme d'optimisation**

# Larges champs d'exploration algorithmique

$$\mathbb{M}_D : \mathbf{g} = \mathbf{H}\mathbf{f} + \epsilon$$

$\mathbf{g}$  : mesures de l'instrument

$\mathbf{H}$  : modèle d'acquisition

$\mathbf{f}$  : objet à reconstruire

$\epsilon$  : incertitudes sur  $\mathbf{g}$  et  $\mathbf{H}$

## Approches bayésiennes

$$\text{Loi a posteriori : } p(\mathbf{f}|\mathbf{g}) = \frac{p(\mathbf{g}|\mathbf{f}; \mathbb{M}_D)p(\mathbf{f}; \mathbb{M}_f)}{p(\mathbf{g})}$$

- 1 Choix de l'estimateur
  - MAP
  - EAP
- 2 Choix du modèle de l'instrument  $\mathbf{H}$
- 3 Choix du modèle des incertitudes  $\epsilon$
- 4 Choix du modèle de l'objet  $\mathbb{M}_f$
- 5 Choix de l'algorithme d'optimisation

# Larges champs d'exploration algorithmique

$$\mathbb{M}_D : \mathbf{g} = \mathbf{H}\mathbf{f} + \epsilon$$

$\mathbf{g}$  : mesures de l'instrument

$\mathbf{H}$  : modèle d'acquisition

$\mathbf{f}$  : objet à reconstruire

$\epsilon$  : incertitudes sur  $\mathbf{g}$  et  $\mathbf{H}$

## Approches bayésiennes

$$\text{Loi a posteriori : } p(\mathbf{f}|\mathbf{g}) = \frac{p(\mathbf{g}|\mathbf{f}; \mathbb{M}_D)p(\mathbf{f}; \mathbb{M}_f)}{p(\mathbf{g})}$$

- 1 Choix de l'estimateur
  - MAP
  - EAP
- 2 Choix du modèle de l'instrument  $\mathbf{H}$
- 3 Choix du modèle des incertitudes  $\epsilon$ 
  - *Lois normale, de student-t* [C<sub>2017d</sub>] ...
  - *Error splitting model* [R<sub>2019b</sub>]
- 4 Choix du modèle de l'objet  $\mathbb{M}_f$
- 5 Choix de l'algorithme d'optimisation

# Larges champs d'exploration algorithmique

$$\mathbb{M}_D : \mathbf{g} = \mathbf{H}\mathbf{f} + \epsilon$$

$\mathbf{g}$  : mesures de l'instrument

$\mathbf{H}$  : modèle d'acquisition

$\mathbf{f}$  : objet à reconstruire

$\epsilon$  : incertitudes sur  $\mathbf{g}$  et  $\mathbf{H}$

## Approches bayésiennes

$$\text{Loi a posteriori : } p(\mathbf{f}|\mathbf{g}) = \frac{p(\mathbf{g}|\mathbf{f}; \mathbb{M}_D)p(\mathbf{f}; \mathbb{M}_f)}{p(\mathbf{g})}$$

- 1 Choix de l'estimateur
  - MAP
  - EAP
- 2 Choix du modèle de l'instrument  $\mathbf{H}$
- 3 Choix du modèle des incertitudes  $\epsilon$ 
  - *Lois normale, de student-t* [C<sub>2017d</sub>] ...
  - *Error splitting model* [R<sub>2019b</sub>]
- 4 Choix du modèle de l'objet  $\mathbb{M}_f$
- 5 Choix de l'algorithme d'optimisation



# Larges champs d'exploration algorithmique

$$\mathbb{M}_D : \mathbf{g} = \mathbf{H}\mathbf{f} + \epsilon$$

$\mathbf{g}$  : mesures de l'instrument

$\mathbf{H}$  : modèle d'acquisition

$\mathbf{f}$  : objet à reconstruire

$\epsilon$  : incertitudes sur  $\mathbf{g}$  et  $\mathbf{H}$

## Approches bayésiennes

$$\text{Loi a posteriori : } p(\mathbf{f}|\mathbf{g}) = \frac{p(\mathbf{g}|\mathbf{f}; \mathbb{M}_D)p(\mathbf{f}; \mathbb{M}_f)}{p(\mathbf{g})}$$

- 1 Choix de l'estimateur
  - MAP
  - EAP
- 2 Choix du modèle de l'instrument  $\mathbf{H}$
- 3 Choix du modèle des incertitudes  $\epsilon$ 
  - *Lois normale, de student-t* [C<sub>2017d</sub>] ...
  - *Error splitting model* [R<sub>2019b</sub>]
- 4 Choix du modèle de l'objet  $\mathbb{M}_f$ 
  - *Modèle de Gauss Markov Potts* [R<sub>2017b</sub>]
  - *Représentation par ondelettes* [R<sub>2018b</sub>]
- 5 Choix de l'algorithme d'optimisation

# Larges champs d'exploration algorithmique

$$\mathbb{M}_D : \mathbf{g} = \mathbf{H}\mathbf{f} + \epsilon$$

$\mathbf{g}$  : mesures de l'instrument

$\mathbf{H}$  : modèle d'acquisition

$\mathbf{f}$  : objet à reconstruire

$\epsilon$  : incertitudes sur  $\mathbf{g}$  et  $\mathbf{H}$

## Approches bayésiennes

$$\text{Loi a posteriori : } p(\mathbf{f}|\mathbf{g}) = \frac{p(\mathbf{g}|\mathbf{f}; \mathbb{M}_D)p(\mathbf{f}; \mathbb{M}_f)}{p(\mathbf{g})}$$

- 1 Choix de l'estimateur
  - MAP
  - EAP
- 2 Choix du modèle de l'instrument  $\mathbf{H}$
- 3 Choix du modèle des incertitudes  $\epsilon$ 
  - *Lois normale, de student-t* [C<sub>2017d</sub>] ...
  - *Error splitting model* [R<sub>2019b</sub>]
- 4 Choix du modèle de l'objet  $\mathbb{M}_f$ 
  - *Modèle de Gauss Markov Potts* [R<sub>2017b</sub>]
  - *Représentation par ondelettes* [R<sub>2018b</sub>]
- 5 Choix de l'algorithme d'optimisation
  - *Descente de gradient simple ou conjuguée*
  - *ADMM, split Bergman...*

# Larges champs d'exploration algorithmique

$$\mathbb{M}_D : \mathbf{g} = \mathbf{H}\mathbf{f} + \boldsymbol{\epsilon}$$

$\mathbf{g}$  : mesures de l'instrument

$\mathbf{H}$  : modèle d'acquisition

$\mathbf{f}$  : objet à reconstruire

$\boldsymbol{\epsilon}$  : incertitudes sur  $\mathbf{g}$  et  $\mathbf{H}$

## Moindre carré reg. quadratique

$$J(\mathbf{f}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 + \lambda\|\mathbf{D}\mathbf{f}\|^2$$

$$\mathbf{f}^{(n+1)} = \mathbf{f}^{(n)} - \alpha\nabla J(\mathbf{f}^{(n)})$$

## Approches bayésiennes

$$\text{Loi a posteriori : } p(\mathbf{f}|\mathbf{g}) = \frac{p(\mathbf{g}|\mathbf{f}; \mathbb{M}_D)p(\mathbf{f}; \mathbb{M}_f)}{p(\mathbf{g})}$$

- 1 Choix de l'estimateur
  - MAP
  - EAP
- 2 Choix du modèle de l'instrument  $\mathbf{H}$
- 3 Choix du modèle des incertitudes  $\boldsymbol{\epsilon}$ 
  - *Lois normale, de student-t* [C<sub>2017d</sub>] ...
  - *Error splitting model* [R<sub>2019b</sub>]
- 4 Choix du modèle de l'objet  $\mathbb{M}_f$ 
  - *Modèle de Gauss Markov Potts* [R<sub>2017b</sub>]
  - *Représentation par ondelettes* [R<sub>2018b</sub>]
- 5 Choix de l'algorithme d'optimisation
  - *Descente de gradient simple ou conjuguée*
  - *ADMM, split Bergman...*

# Larges champs d'exploration algorithmique

$$\mathbb{M}_D : \mathbf{g} = \mathbf{H}\mathbf{f} + \boldsymbol{\epsilon}$$

$\mathbf{g}$  : mesures de l'instrument

$\mathbf{H}$  : modèle d'acquisition

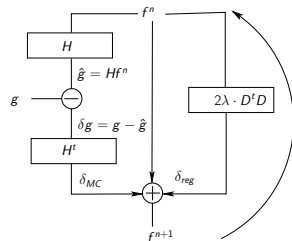
$\mathbf{f}$  : objet à reconstruire

$\boldsymbol{\epsilon}$  : incertitudes sur  $\mathbf{g}$  et  $\mathbf{H}$

## Moindre carré reg. quadratique

$$J(\mathbf{f}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 + \lambda\|\mathbf{D}\mathbf{f}\|^2$$

$$\mathbf{f}^{(n+1)} = \mathbf{f}^{(n)} - \alpha\nabla J(\mathbf{f}^{(n)})$$



## Approches bayésiennes

$$\text{Loi a posteriori} : p(\mathbf{f}|\mathbf{g}) = \frac{p(\mathbf{g}|\mathbf{f}; \mathbb{M}_D)p(\mathbf{f}; \mathbb{M}_f)}{p(\mathbf{g})}$$

- 1 Choix de l'estimateur
  - MAP
  - EAP
- 2 Choix du modèle de l'instrument  $\mathbf{H}$
- 3 Choix du modèle des incertitudes  $\boldsymbol{\epsilon}$ 
  - *Lois normale, de student-t* [C<sub>2017d</sub>] ...
  - *Error splitting model* [R<sub>2019b</sub>]
- 4 Choix du modèle de l'objet  $\mathbb{M}_f$ 
  - *Modèle de Gauss Markov Potts* [R<sub>2017b</sub>]
  - *Représentation par ondelettes* [R<sub>2018b</sub>]
- 5 Choix de l'algorithme d'optimisation
  - *Descente de gradient simple ou conjuguée*
  - *ADMM, split Bergman...*

# Larges champs d'exploration algorithmique

$$\mathbb{M}_D : \mathbf{g} = \mathbf{H}\mathbf{f} + \boldsymbol{\epsilon}$$

$\mathbf{g}$  : mesures de l'instrument

$\mathbf{H}$  : modèle d'acquisition

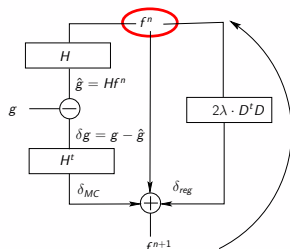
$\mathbf{f}$  : objet à reconstruire

$\boldsymbol{\epsilon}$  : incertitudes sur  $\mathbf{g}$  et  $\mathbf{H}$

## Moindre carré reg. quadratique

$$J(\mathbf{f}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 + \lambda\|\mathbf{D}\mathbf{f}\|^2$$

$$\mathbf{f}^{(n+1)} = \mathbf{f}^{(n)} - \alpha \nabla J(\mathbf{f}^{(n)})$$



## Approches bayésiennes

$$\text{Loi a posteriori} : p(\mathbf{f}|\mathbf{g}) = \frac{p(\mathbf{g}|\mathbf{f}; \mathbb{M}_D)p(\mathbf{f}; \mathbb{M}_f)}{p(\mathbf{g})}$$

- 1 Choix de l'estimateur
  - MAP
  - EAP
- 2 Choix du modèle de l'instrument  $\mathbf{H}$
- 3 Choix du modèle des incertitudes  $\boldsymbol{\epsilon}$ 
  - *Lois normale, de student-t* [C<sub>2017d</sub>] ...
  - *Error splitting model* [R<sub>2019b</sub>]
- 4 Choix du modèle de l'objet  $\mathbb{M}_f$ 
  - *Modèle de Gauss Markov Potts* [R<sub>2017b</sub>]
  - *Représentation par ondelettes* [R<sub>2018b</sub>]
- 5 Choix de l'algorithme d'optimisation
  - *Descente de gradient simple ou conjuguée*
  - *ADMM, split Bergman...*

# Larges champs d'exploration algorithmique

$$\mathbb{M}_D : \mathbf{g} = \mathbf{H}\mathbf{f} + \epsilon$$

$\mathbf{g}$  : mesures de l'instrument

$\mathbf{H}$  : modèle d'acquisition

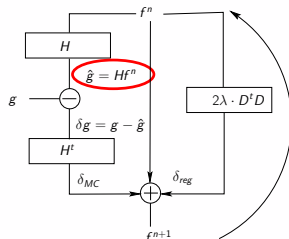
$\mathbf{f}$  : objet à reconstruire

$\epsilon$  : incertitudes sur  $\mathbf{g}$  et  $\mathbf{H}$

## Moindre carré reg. quadratique

$$J(\mathbf{f}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 + \lambda\|\mathbf{D}\mathbf{f}\|^2$$

$$\mathbf{f}^{(n+1)} = \mathbf{f}^{(n)} - \alpha\nabla J(\mathbf{f}^{(n)})$$



## Approches bayésiennes

$$\text{Loi a posteriori} : p(\mathbf{f}|\mathbf{g}) = \frac{p(\mathbf{g}|\mathbf{f}; \mathbb{M}_D)p(\mathbf{f}; \mathbb{M}_f)}{p(\mathbf{g})}$$

- 1 Choix de l'estimateur
  - MAP
  - EAP
- 2 Choix du modèle de l'instrument  $\mathbf{H}$
- 3 Choix du modèle des incertitudes  $\epsilon$ 
  - *Lois normale, de student-t* [C<sub>2017d</sub>] ...
  - *Error splitting model* [R<sub>2019b</sub>]
- 4 Choix du modèle de l'objet  $\mathbb{M}_f$ 
  - *Modèle de Gauss Markov Potts* [R<sub>2017b</sub>]
  - *Représentation par ondelettes* [R<sub>2018b</sub>]
- 5 Choix de l'algorithme d'optimisation
  - *Descente de gradient simple ou conjuguée*
  - *ADMM, split Bergman...*

# Larges champs d'exploration algorithmique

$$\mathbb{M}_D : \mathbf{g} = \mathbf{H}\mathbf{f} + \boldsymbol{\epsilon}$$

$\mathbf{g}$  : mesures de l'instrument

$\mathbf{H}$  : modèle d'acquisition

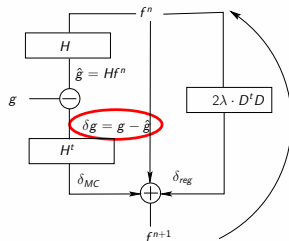
$\mathbf{f}$  : objet à reconstruire

$\boldsymbol{\epsilon}$  : incertitudes sur  $\mathbf{g}$  et  $\mathbf{H}$

## Moindre carré reg. quadratique

$$J(\mathbf{f}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 + \lambda\|\mathbf{D}\mathbf{f}\|^2$$

$$\mathbf{f}^{(n+1)} = \mathbf{f}^{(n)} - \alpha\nabla J(\mathbf{f}^{(n)})$$



## Approches bayésiennes

$$\text{Loi a posteriori} : p(\mathbf{f}|\mathbf{g}) = \frac{p(\mathbf{g}|\mathbf{f}; \mathbb{M}_D)p(\mathbf{f}; \mathbb{M}_f)}{p(\mathbf{g})}$$

- 1 Choix de l'estimateur
  - MAP
  - EAP
- 2 Choix du modèle de l'instrument  $\mathbf{H}$
- 3 Choix du modèle des incertitudes  $\boldsymbol{\epsilon}$ 
  - *Lois normale, de student-t* [C<sub>2017d</sub>] ...
  - *Error splitting model* [R<sub>2019b</sub>]
- 4 Choix du modèle de l'objet  $\mathbb{M}_f$ 
  - *Modèle de Gauss Markov Potts* [R<sub>2017b</sub>]
  - *Représentation par ondelettes* [R<sub>2018b</sub>]
- 5 Choix de l'algorithme d'optimisation
  - *Descente de gradient simple ou conjuguée*
  - *ADMM, split Bergman...*

# Larges champs d'exploration algorithmique

$$\mathbb{M}_D : \mathbf{g} = \mathbf{H}\mathbf{f} + \boldsymbol{\epsilon}$$

$\mathbf{g}$  : mesures de l'instrument

$\mathbf{H}$  : modèle d'acquisition

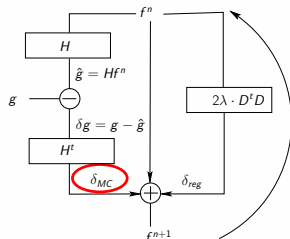
$\mathbf{f}$  : objet à reconstruire

$\boldsymbol{\epsilon}$  : incertitudes sur  $\mathbf{g}$  et  $\mathbf{H}$

## Moindre carré reg. quadratique

$$J(\mathbf{f}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 + \lambda\|\mathbf{D}\mathbf{f}\|^2$$

$$\mathbf{f}^{(n+1)} = \mathbf{f}^{(n)} - \alpha \nabla J(\mathbf{f}^{(n)})$$



## Approches bayésiennes

$$\text{Loi a posteriori : } p(\mathbf{f}|\mathbf{g}) = \frac{p(\mathbf{g}|\mathbf{f}; \mathbb{M}_D)p(\mathbf{f}; \mathbb{M}_f)}{p(\mathbf{g})}$$

- 1 Choix de l'estimateur
  - MAP
  - EAP
- 2 Choix du modèle de l'instrument  $\mathbf{H}$
- 3 Choix du modèle des incertitudes  $\boldsymbol{\epsilon}$ 
  - *Lois normale, de student-t* [C<sub>2017d</sub>] ...
  - *Error splitting model* [R<sub>2019b</sub>]
- 4 Choix du modèle de l'objet  $\mathbb{M}_f$ 
  - *Modèle de Gauss Markov Potts* [R<sub>2017b</sub>]
  - *Représentation par ondelettes* [R<sub>2018b</sub>]
- 5 Choix de l'algorithme d'optimisation
  - *Descente de gradient simple ou conjuguée*
  - *ADMM, split Bergman...*



# Larges champs d'exploration algorithmique

$$\mathbb{M}_D : \mathbf{g} = \mathbf{H}\mathbf{f} + \epsilon$$

$\mathbf{g}$  : mesures de l'instrument

$\mathbf{H}$  : modèle d'acquisition

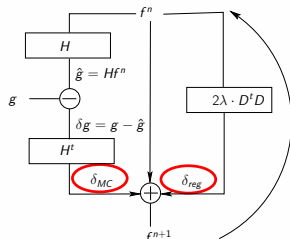
$\mathbf{f}$  : objet à reconstruire

$\epsilon$  : incertitudes sur  $\mathbf{g}$  et  $\mathbf{H}$

## Moindre carré reg. quadratique

$$J(\mathbf{f}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 + \lambda\|\mathbf{D}\mathbf{f}\|^2$$

$$\mathbf{f}^{(n+1)} = \mathbf{f}^{(n)} - \alpha\nabla J(\mathbf{f}^{(n)})$$



## Approches bayésiennes

$$\text{Loi a posteriori : } p(\mathbf{f}|\mathbf{g}) = \frac{p(\mathbf{g}|\mathbf{f}; \mathbb{M}_D)p(\mathbf{f}; \mathbb{M}_f)}{p(\mathbf{g})}$$

- Choix de l'estimateur
  - MAP
  - EAP
- Choix du modèle de l'instrument  $\mathbf{H}$
- Choix du modèle des incertitudes  $\epsilon$ 
  - Lois normale, de student-t [C<sub>2017d</sub>] ...
  - Error splitting model [R<sub>2019b</sub>]
- Choix du modèle de l'objet  $\mathbb{M}_f$ 
  - Modèle de Gauss Markov Potts [R<sub>2017b</sub>]
  - Représentation par ondelettes [R<sub>2018b</sub>]
- Choix de l'algorithme d'optimisation
  - Descente de gradient simple ou conjuguée
  - ADMM, split Bergman...

# Larges champs d'exploration algorithmique

$$\mathbb{M}_D : \mathbf{g} = \mathbf{H}\mathbf{f} + \boldsymbol{\epsilon}$$

$\mathbf{g}$  : mesures de l'instrument

$\mathbf{H}$  : modèle d'acquisition

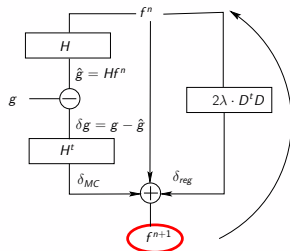
$\mathbf{f}$  : objet à reconstruire

$\boldsymbol{\epsilon}$  : incertitudes sur  $\mathbf{g}$  et  $\mathbf{H}$

## Moindre carré reg. quadratique

$$J(\mathbf{f}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 + \lambda\|\mathbf{D}\mathbf{f}\|^2$$

$$\mathbf{f}^{(n+1)} = \mathbf{f}^{(n)} - \alpha \nabla J(\mathbf{f}^{(n)})$$



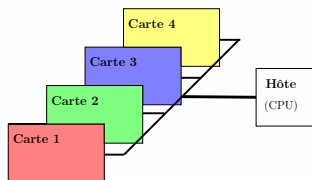
## Approches bayésiennes

$$\text{Loi a posteriori} : p(\mathbf{f}|\mathbf{g}) = \frac{p(\mathbf{g}|\mathbf{f}; \mathbb{M}_D)p(\mathbf{f}; \mathbb{M}_f)}{p(\mathbf{g})}$$

- 1 Choix de l'estimateur
  - MAP
  - EAP
- 2 Choix du modèle de l'instrument  $\mathbf{H}$
- 3 Choix du modèle des incertitudes  $\boldsymbol{\epsilon}$ 
  - *Lois normale, de student-t* [C<sub>2017d</sub>] ...
  - *Error splitting model* [R<sub>2019b</sub>]
- 4 Choix du modèle de l'objet  $\mathbb{M}_f$ 
  - *Modèle de Gauss Markov Potts* [R<sub>2017b</sub>]
  - *Représentation par ondelettes* [R<sub>2018b</sub>]
- 5 Choix de l'algorithme d'optimisation
  - *Descente de gradient simple ou conjuguée*
  - *ADMM, split Bergman...*

# Architectures de calcul étudiées

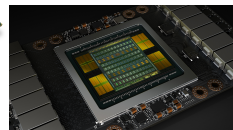
## Plateforme de calcul avec cartes accélératrices



(a) plateforme de calcul



(b) carte PCIe



(c) puce de calcul

## Puces accélératrices de calcul

- **GPU** : processeur massivement parallèle
- **FPGA** : conception d'une architecture sur mesure

⇒ *Adaptées aux systèmes embarqués, PC ou supercalculateur HPC*

# Larges champs d'exploration architecturale

## GPU : algorithmes massivement parallèles

- **Parallélisation**

- #1 à grains fins sur les [100 - 5000] coeurs de calcul

- #2 à grains épais sur les [1-16] cartes accélératrices

- **Murs mémoire pour l'accès aux données**

- #1 coeurs de calcul ↔ mémoire DDR [1 - 32] Go

- #2 cartes accélératrices ↔ mémoire hôte [0.1 - 1] To

- **Boîte à outils d'accélération** (texture 2D/3D, tensor cores, *half float*...)

# Larges champs d'exploration architecturale

## GPU : algorithmes massivement parallèles

- **Parallélisation**

  - #1 à grains fins sur les [100 - 5000] coeurs de calcul

  - #2 à grains épais sur les [1-16] cartes accélératrices

- **Murs mémoire pour l'accès aux données**

  - #1 coeurs de calcul ↔ mémoire DDR [1 - 32] Go

  - #2 cartes accélératrices ↔ mémoire hôte [0.1 - 1] To

- **Boîte à outils d'accélération** (texture 2D/3D, tensor cores, *half float*...)

## FPGA : outils de synthèse de haut niveau (HLS)

- **Architecture alternative au GPU**

  - #1 Architecture basse consommation

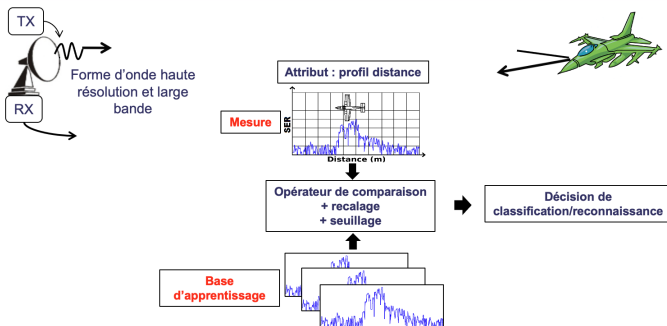
  - #2 Parallélisme de pipeline

- **Guide/méthodologie d'utilisation des outils HLS**

## Reconnaissance radar

2010-13

Thomas Boulay - CIFRE avec Thales Air Systems

*Classification de signatures radars accélérée sur GPUs*[R<sub>2013</sub>]

## Imagerie acoustique

2010-13

Ning Chu

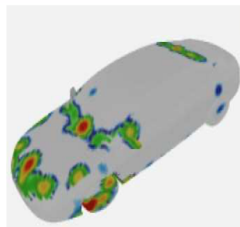
## Reconnaissance radar

2010-13      Thomas Boulay

## Imagerie acoustique

2010-13      **Ning Chu** - *Bourse China Scholarship Council*

*Approches bayésiennes pour la localisation de sources acoustiques*

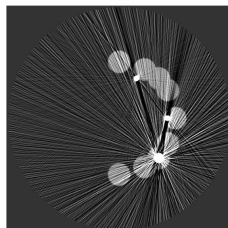
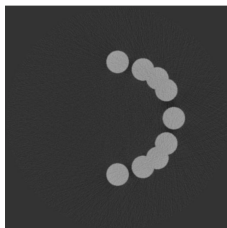


[R<sub>2014a</sub>, C<sub>2014b</sub>]

## Reconstruction tomographique en imagerie à rayons X

2011-14 Long Chen - *CIFRE avec Carestream Dental*

Réduction d'artefacts métalliques pour l'imagerie dentaire

[C<sub>2013d</sub>, C<sub>2014d</sub>]

2014-17 Li Wang

2016-19 Camille Chapdelaine

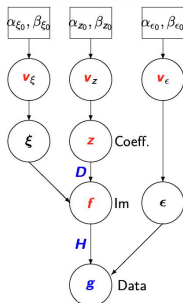


## Reconstruction tomographique en imagerie à rayons X

2011-14 Long Chen

2014-17 **Li Wang** - *Bourse China Scholarship Council*

Approches bayésiennes hiérarchiques pour la reconstruction

[R<sub>2017a</sub>, R<sub>2018b</sub>]

2016-19 Camille Chapdelaine

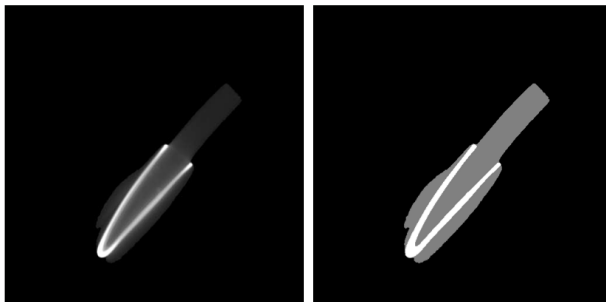
## Reconstruction tomographique en imagerie à rayons X

2011-14 Long Chen

2014-17 Li Wang

2016-19 **Camille Chapdelaine** - *CIFRE avec SAFRAN*

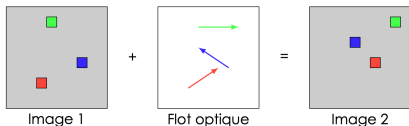
Contrôle Non Destructif (CND) de pièces aéronautiques



[R<sub>2017b</sub>, R<sub>2019b</sub>]

## Parallélisation GPU

2018-21

**Mickael Seznec** - *CIFRE avec Thales TRT**Démarche  $A^3$  pour le calcul haute performance embarqué*

*carte GPU  
pour système embarqué*

[ $C_{2018b}, C_{2020b}$ ]

2019-22

Nicolas Monnier

## Synthèse de haut niveau sur FPGA

2016-19

Maxime Martelli

2019-22

Daouda Diakite

## Parallélisation GPU

2018-21 Mickael Seznec

2019-22 **Nicolas Monnier** - *Bourse Région IdF + coll. Atos Bull*

*ExaSKA : Parallélisation d'algorithmes TdSI en radioastronomie*



carte GPU  
pour système HPC

## Synthèse de haut niveau sur FPGA

2016-19 Maxime Martelli

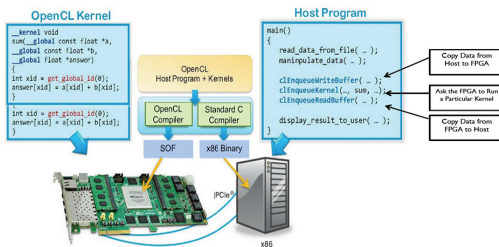
2019-22 Daouda Diakite

## Parallélisation GPU

2018-21 Mickael Seznec

2019-22 Nicolas Monnier

## Synthèse de haut niveau sur FPGA

2016-19 **Maxime Martelli** - *CIFRE avec Thales DMS**Approche haut niveau pour l'accélération de la simulation radar*[R<sub>2018a</sub>]

2019-22 Daouda Diakite

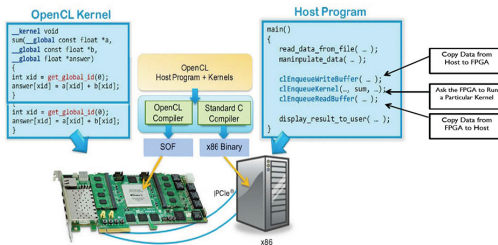
## Parallélisation GPU

2018-21	Mickael Seznec
2019-22	Nicolas Monnier

## Synthèse de haut niveau sur FPGA

2016-19	Maxime Martelli
2019-22	<b>Daouda Diakite</b> - <i>Bourse ED STIC</i>

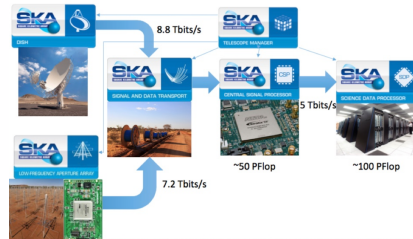
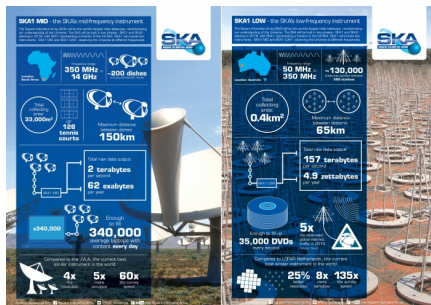
*Démarche A<sup>3</sup> appliquée à la tomographie et à la radioastronomie*

[C<sub>2020a</sub>]

# Projets collaboratifs

## Radiotélescope SKA

### ● Pipeline de formation des images hyperspectrales



- 2015-... Lagrange (A. Ferrari) - ANR Magellan puis DeconvSKA
- 2018-19 IETR/Lagrange/Obs Paris/Atos Bull - PEPS **SKALLAS** (porteur)
- 2021-25 IETR/Lagrange/Obs Nançay/IRISA - ANR **DARK-ERA** (porteur)

## Projets collaboratifs

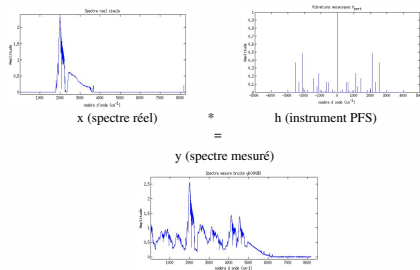
## Traitements de données spectrales

- Correction de vibrations mécaniques par déconvolution 1D

2010-13 GEOPS (F. Schmidt)



Instrument PFS (Planetary Fourier Spectrum)



[R<sub>2014b</sub>]

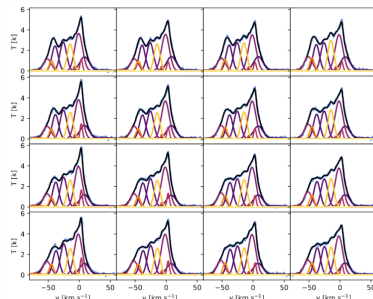
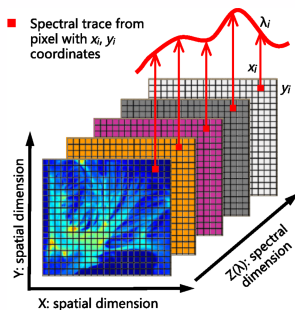
- Démélange de gaz interstellaires



# Projets collaboratifs

## Traitements de données spectrales

- Correction de vibrations mécaniques par déconvolution 1D
- Démélange de gaz interstellaires



2017-18 CEA-Irfu/LIP6 - CNRS HyperStars (M.A. Miville-Deschênes)

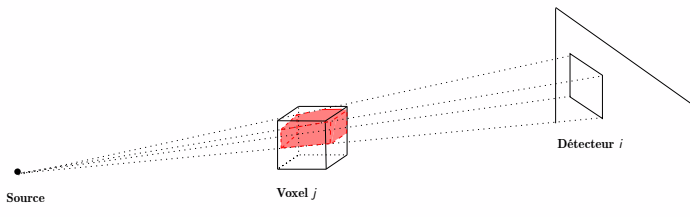
2018-21 Parcours recherche de J. Besson  
Méthode ROHSA [R<sub>2019a</sub>] parallélisée sur GPU

- 1 Parcours depuis le doctorat
- 2 Adéquation Algorithme Architecture ( $A^3$ )
  - Problématique et positionnement
  - $A^3$  au sein du GPI
  - Encadrement de thèses et collaborations
- 3 Calcul intensif sur GPU/FPGA
  - Paire de projection/rétroprojection  $\mathcal{P}/\mathcal{R}$
  - Opérateur de convolution  $\mathcal{C}$
- 4 Calcul distribué sur serveur multiGPU
  - Reconstruction tomographique
  - DeconvSKA : Déconvolution en radioastronomie
- 5 Perspectives
  - Une méthodologie  $A^3$  à échelle humaine
  - SKA, un défi  $A^3$  à grande échelle

Opérateurs  $\mathcal{P}$  de projection et  $\mathcal{R}$  de rétroprojection

## Algorithme itératif

$$\mathbf{f}^{(n+1)} = \mathbf{f}^{(n)} + \alpha \mathbf{H}^t(\mathbf{g} - \mathbf{H}\mathbf{f}) + \dots$$

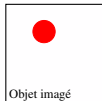
 $\mathbf{H}_{ij}$  : contribution du voxel  $j$  dans le détecteur  $i$  $\mathbf{H}$  de grande dimension

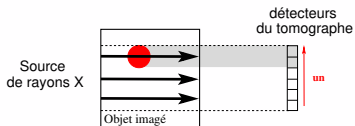
$\mathbf{f}$  de taille  $2048^3 \implies \mathbf{H}$  de taille 256 Exa octets

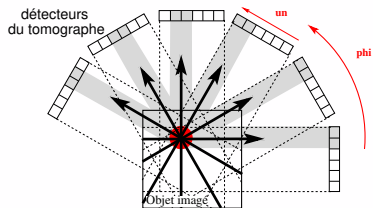
$\implies$  Calcul de  $\mathbf{H}$  et de  $\mathbf{H}^t$  par une paire  $\mathcal{P}/\mathcal{R}$

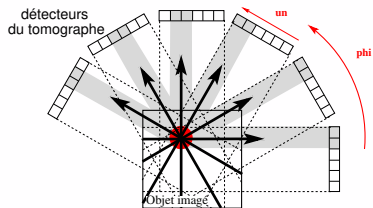
$\mathcal{P}/\mathcal{R}$  : boucles naturelles de parallélisation

Opérateur de projection  $\mathcal{P}$



$\mathcal{P}/\mathcal{R}$  : boucles naturelles de parallélisationOpérateur de projection  $\mathcal{P}$ 

$\mathcal{P}/\mathcal{R}$  : boucles naturelles de parallélisationOpérateur de projection  $\mathcal{P}$ 

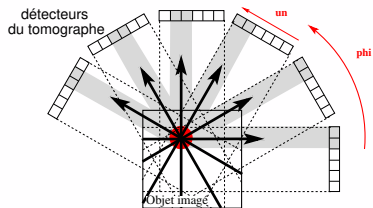
$\mathcal{P}/\mathcal{R}$  : boucles naturelles de parallélisationOpérateur de projection  $\mathcal{P}$ 

```
for un, phi do
```

```
  ...
```

```
   $\mathcal{P}(\text{un}, \text{phi}) = \dots$ 
```

```
end for
```

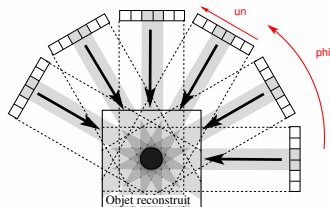
$\mathcal{P}/\mathcal{R}$  : boucles naturelles de parallélisationOpérateur de projection  $\mathcal{P}$ 

```
for un, phi do
```

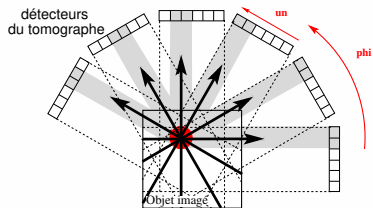
```
...
```

```
 $\mathcal{P}(\text{un}, \text{phi}) = \dots$ 
```

```
end for
```

Opérateur de rétroprojection  $\mathcal{R}$ 

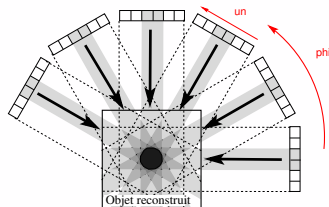


$\mathcal{P}/\mathcal{R}$  : boucles naturelles de parallélisationOpérateur de projection  $\mathcal{P}$ 

```

for un, phi do
  ...
   $\mathcal{P}(\text{un}, \text{phi}) = \dots$ 
end for

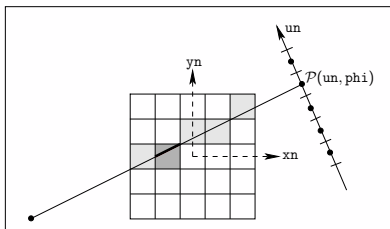
```

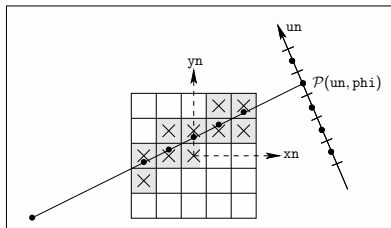
Opérateur de rétroprojection  $\mathcal{R}$ 

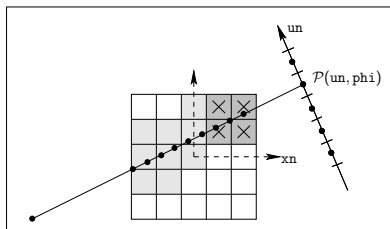
```

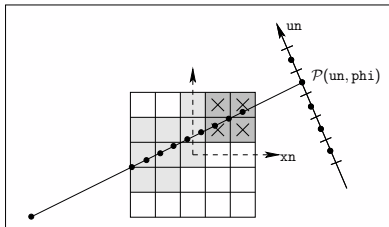
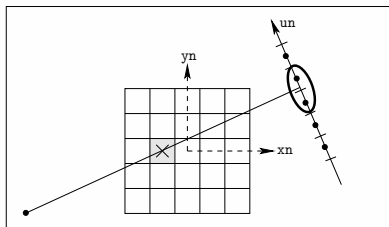
for xn, yn do
  ...
   $\mathcal{R}(\text{xn}, \text{yn}) = \dots$ 
end for

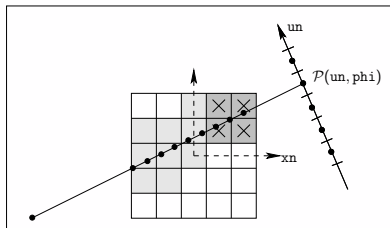
```

Modèle *ray-driven*  $\mathcal{P}_{Siddon}$ 

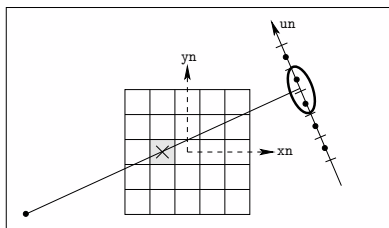
Modèle *ray-driven*  $\mathcal{P}_{Joseph}$ 

Modèle *ray-driven*  $\mathcal{P}_{Regular}$ 

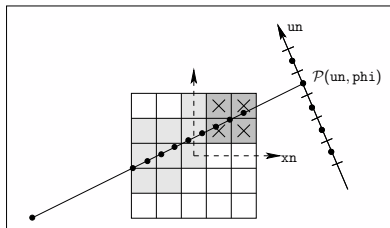
Modèle *ray-driven*  $\mathcal{P}_{Regular}$ Modèle *Voxel-Driven*  $\mathcal{R}_{VDL}$ 

Modèle *ray-driven*  $\mathcal{P}_{Regular}$ 

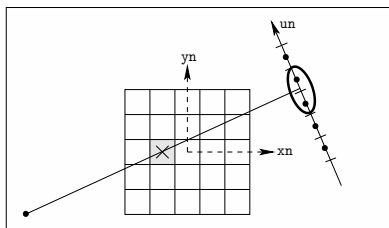
$$\mathcal{P}(f) = H_{\mathcal{P}} f$$

Modèle *Voxel-Driven*  $\mathcal{R}_{VDL}$ 

$$\mathcal{R}(g) = H_{\mathcal{R}}^t g$$

Modèle *ray-driven*  $\mathcal{P}_{Regular}$ 

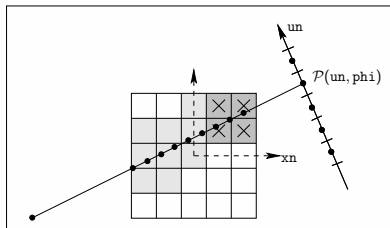
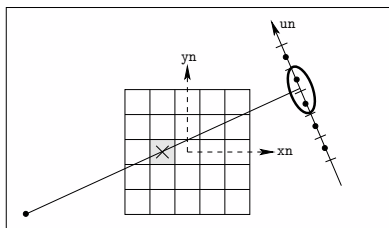
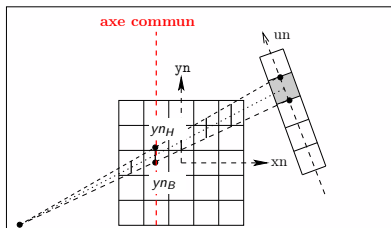
$$\mathcal{P}(\mathbf{f}) = \mathbf{H}_{\mathcal{P}} \mathbf{f}$$

Modèle *Voxel-Driven*  $\mathcal{R}_{VDL}$ 

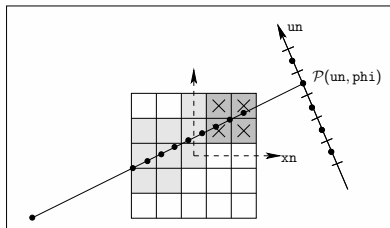
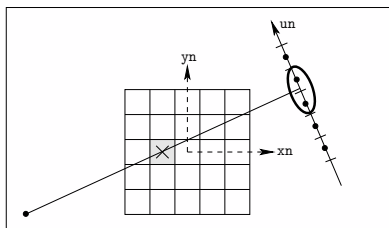
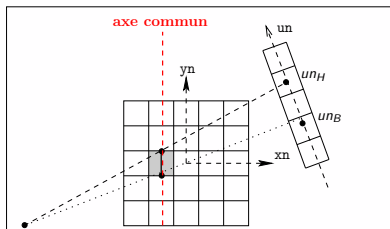
$$\mathcal{R}(\mathbf{g}) = \mathbf{H}_{\mathcal{R}}^t \mathbf{g}$$

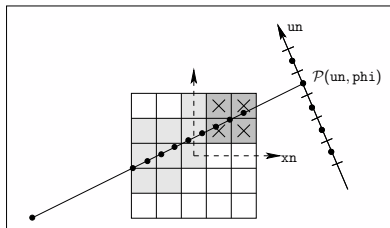
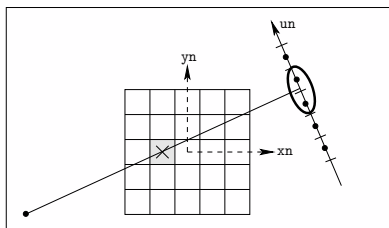
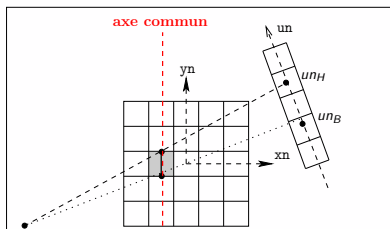
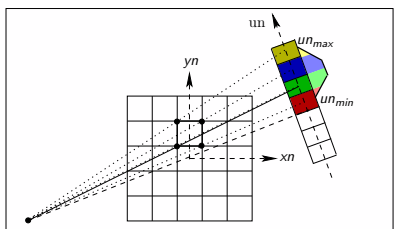
La paire  $\mathcal{P}_{Regular}/\mathcal{R}_{VDL}$  est **non duale**

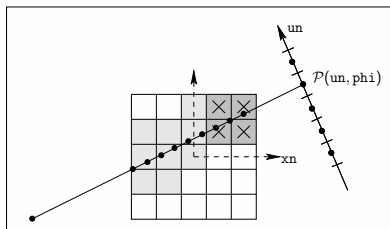
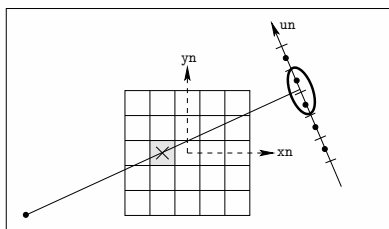
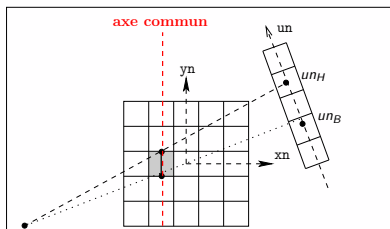
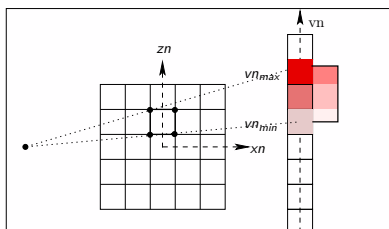
$$\mathcal{P}/\mathcal{R} \text{ est une paire duale} \quad \equiv \quad \forall \mathbf{f}, \forall \mathbf{g}, \langle \mathbf{g} \cdot \mathcal{P}(\mathbf{f}) \rangle = \langle \mathcal{R}(\mathbf{g}) \cdot \mathbf{f} \rangle$$

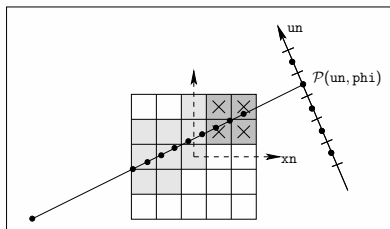
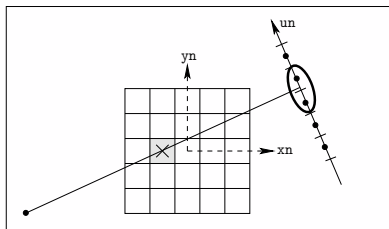
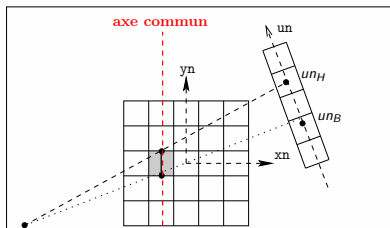
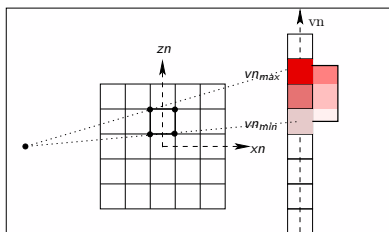
Modèle *ray-driven*  $\mathcal{P}_{Regular}$ Modèle *Voxel-Driven*  $\mathcal{R}_{VDL}$ Modèle *Distance-Driven*  $\mathcal{P}_{DD}$ 



Modèle *ray-driven*  $\mathcal{P}_{Regular}$ Modèle *Voxel-Driven*  $\mathcal{R}_{VDL}$ Modèle *Distance-Driven*  $\mathcal{R}_{DD}$ 

Modèle *ray-driven*  $\mathcal{P}_{Regular}$ Modèle *Voxel-Driven*  $\mathcal{R}_{VDL}$ Modèle *Distance-Driven*  $\mathcal{R}_{DD}$ Modèle *Separable Footprint (transverse)*

Modèle *ray-driven*  $\mathcal{P}_{Regular}$ Modèle *Voxel-Driven*  $\mathcal{R}_{VDL}$ Modèle *Distance-Driven*  $\mathcal{R}_{DD}$ Modèle *Separable Footprint (axial)*

Modèle *ray-driven*  $\mathcal{P}_{Regular}$ Modèle *Voxel-Driven*  $\mathcal{R}_{VDL}$ Modèle *Distance-Driven*  $\mathcal{R}_{DD}$ Modèle *Separable Footprint (axial)*

$\Rightarrow \mathcal{P}_{DD}/\mathcal{R}_{DD}$  et  $\mathcal{P}_{SF}/\mathcal{R}_{SF}$  paires duales

## Contributions (1/2)

## Accélération sur GPU

- Mise en évidence des limitations dûes aux paires non duales
- Parallélisation originale de la paire duale  $\mathcal{P}_{SF}/\mathcal{R}_{SF}$  [ $C_{2018a}, C_{2019a}$ ]
- Comparaison de performances

	TomoGPI (s)	Astra (s)	RTK (s)
$\mathcal{P}_{Regular}$	5.34	6.74 ( $\times 1.3$ )	41.6 ( $\times 7.8$ )
$\mathcal{R}_{VDL}$	2.60	5.49 ( $\times 2.1$ )	5.63 ( $\times 2.2$ )

GPU<sub>volta-Titan</sub> ; jeu de données [ $1024 \times 1024^2$ ;  $1024^3$ ]

## Accélération sur FPGA

- Rétroprojecteur  $\mathcal{R}_{VDL}$  parallélisé sur FPGA avec les outils HLS [ $R_{2018a}$ ]

# Contributions (2/2)

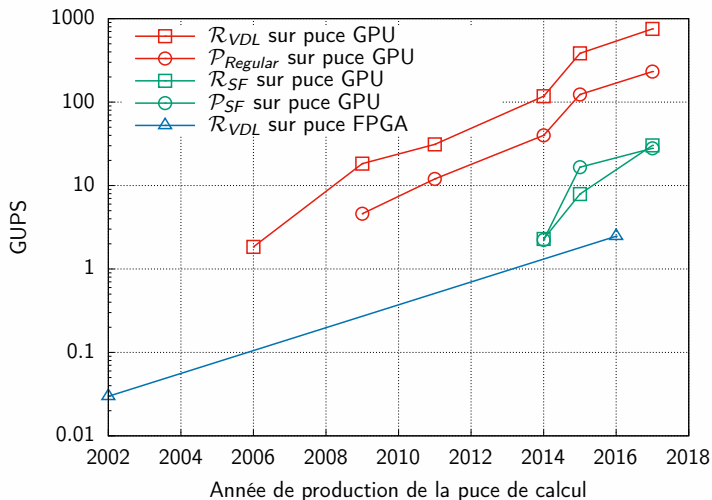
## Logiciels

- TomoGPI : fruit des travaux du GPI depuis 10 ans
- TomoBayes : co-propriété du L2S et de SAFRAN
- Volonté de rendre les codes ouverts

*Laptop démonstrateur  
avec boîtiers externes  
FPGA et GPU*



## GPU versus FPGA

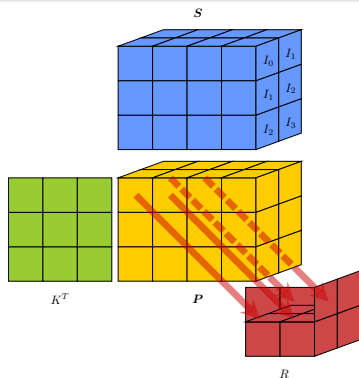
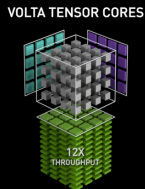
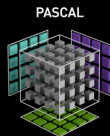
GUPS : Giga mises à jour de voxels ( $\mathcal{R}$ ) ou de rayons ( $\mathcal{P}$ ) par seconde

# Accélération de la convolution 2D sur GPU

Le GPU, une boîte à outils d'accélération [M. Seznec]

- Calcul et stockage en demi-flottant pour la déconvolution [C<sub>2018b</sub>]
- *tensors cores* pour noyaux de taille "moyenne" [C<sub>2020d</sub>]

$$D = \begin{pmatrix} A_{1,0} & A_{1,1} & A_{1,2} & A_{1,3} \\ A_{1,0} & A_{1,1} & A_{1,2} & A_{1,3} \\ A_{2,0} & A_{2,1} & A_{2,2} & A_{2,3} \\ A_{2,0} & A_{2,1} & A_{2,2} & A_{2,3} \\ A_{3,0} & A_{3,1} & A_{3,2} & A_{3,3} \end{pmatrix} + \begin{pmatrix} B_{1,0} & B_{1,1} & B_{1,2} & B_{1,3} \\ B_{1,0} & B_{1,1} & B_{1,2} & B_{1,3} \\ B_{2,0} & B_{2,1} & B_{2,2} & B_{2,3} \\ B_{2,0} & B_{2,1} & B_{2,2} & B_{2,3} \\ B_{3,0} & B_{3,1} & B_{3,2} & B_{3,3} \end{pmatrix} + \begin{pmatrix} C_{0,0} & C_{0,1} & C_{0,2} & C_{0,3} \\ C_{1,0} & C_{1,1} & C_{1,2} & C_{1,3} \\ C_{2,0} & C_{2,1} & C_{2,2} & C_{2,3} \\ C_{3,0} & C_{3,1} & C_{3,2} & C_{3,3} \end{pmatrix}$$

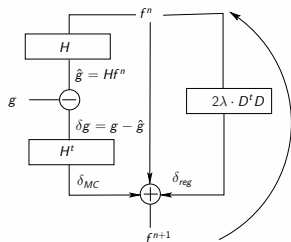


Algorithme *im2tensor*

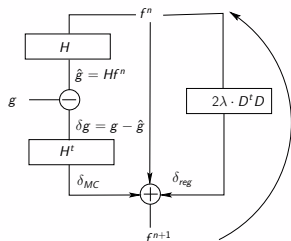


- 1 Parcours depuis le doctorat
- 2 Adéquation Algorithme Architecture ( $A^3$ )
  - Problématique et positionnement
  - $A^3$  au sein du GPI
  - Encadrement de thèses et collaborations
- 3 Calcul intensif sur GPU/FPGA
  - Paire de projection/rétroprojection  $\mathcal{P}/\mathcal{R}$
  - Opérateur de convolution  $\mathcal{C}$
- 4 Calcul distribué sur serveur multiGPU
  - Reconstruction tomographique
  - DeconvSKA : Déconvolution en radioastronomie
- 5 Perspectives
  - Une méthodologie  $A^3$  à échelle humaine
  - SKA, un défi  $A^3$  à grande échelle

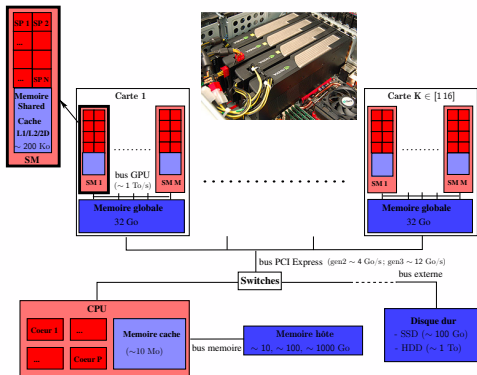
## Boucle itérative



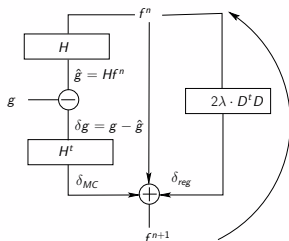
## Boucle itérative



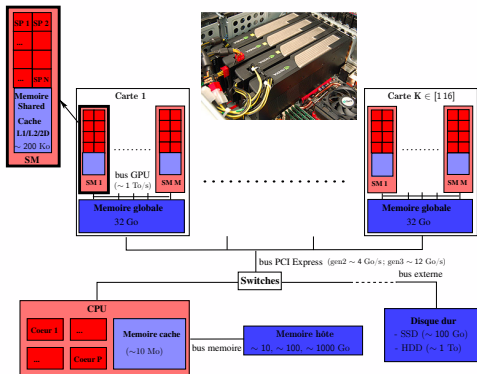
## Le bus PCIe potentiel goulôt d'étranglement



## Boucle itérative



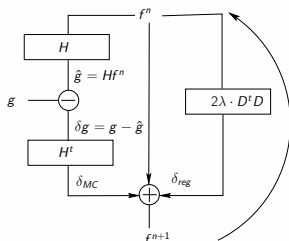
## Le bus PCIe potentiel goulôt d'étranglement



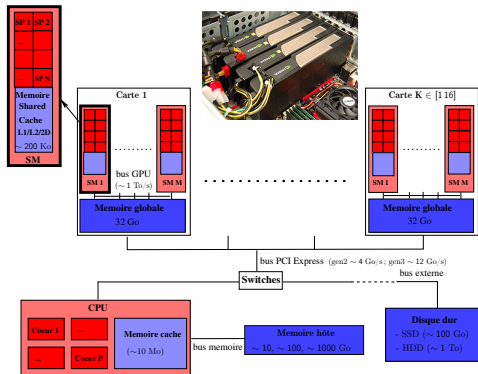
## Stratégies en grande dimension

#1 Recouvrement des temps de transfert mémoire (*streams*)

## Boucle itérative



## Le bus PCIe potentiel goulôt d'étranglement

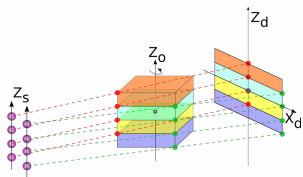


## Stratégies en grande dimension

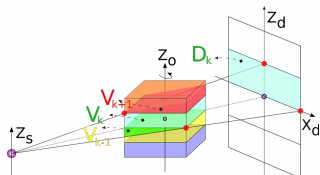
- #1 Recouvrement des temps de transfert mémoire (*streams*)
- #2 Décentralisation des données avec échange de données entre GPUs voisins

Distribution de données en tomographie [ $C_{2020c}$ ]

## Difficulté en géométrie conique



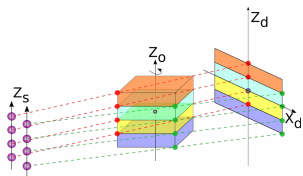
(géométrie parallèle)



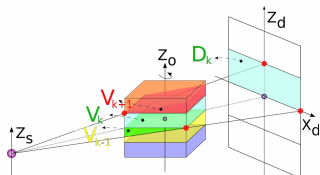
(géométrie conique)

Distribution de données en tomographie [ $C_{2020c}$ ]

## Difficulté en géométrie conique



(géométrie parallèle)

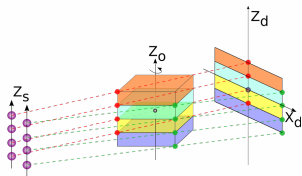


(géométrie conique)

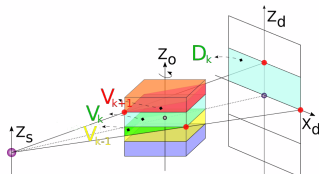
⇒ transfert p2p avant chaque appel à  $\mathcal{P}/\mathcal{R}$

Distribution de données en tomographie [C<sub>2020c</sub>]

## Difficulté en géométrie conique

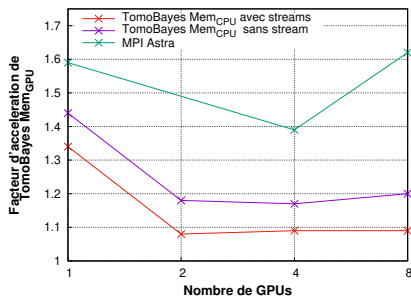


(géométrie parallèle)



(géométrie conique)

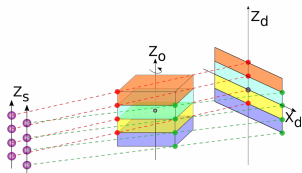
⇒ transfert p2p avant chaque appel à  $\mathcal{P}/\mathcal{R}$

Accélération de HtH sur  $S_{Zay}$  [GPU<sub>V100</sub>]

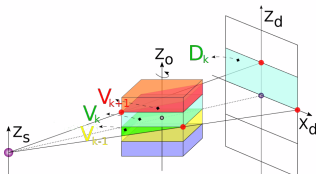


Distribution de données en tomographie [ $C_{2020c}$ ]

## Difficulté en géométrie conique

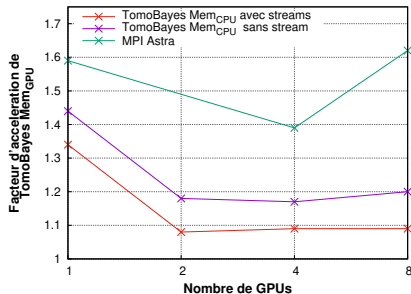


(géométrie parallèle)



(géométrie conique)

⇒ transfert p2p avant chaque appel à  $\mathcal{P}/\mathcal{R}$

Accélération de HtH sur  $S_{zay}$  [GPU $V_{100}$ ]Descente de gradient sur  $S_{zay}$  [GPU $V_{100}$ ]

TomoBayes		MPI Astra
Mem <sub>GPU</sub>	Mem <sub>CPU</sub> ✗	✗
2.55 mn	17.07 mn	3.84
	×6.70	×1.51

# Distribution des données en radioastronomie

## Déconvolution

- Données : Images *dirty*  $\tilde{\mathbf{I}}_\ell$  du ciel à plusieurs longueurs d'onde  $\ell$
- Critère : attache aux données + **reg. spatiale** + **reg. spectrale**

$$J(\mathbf{f}) = \sum_{\ell=1}^L \|\tilde{\mathbf{I}}_\ell - \mathbf{H}_\ell \mathbf{I}_\ell\|_2^2 + \mu_I \sum_{\ell=1}^L \|\mathbf{D}_I \mathbf{I}_\ell\|_2^2 + \mu_S \sum_{i,j=1,1}^{N^2} \|\mathbf{D}_S \mathbf{S}_{i,j}\|_2^2,$$

- Descente :  $\mathbf{f}^{(m+1)} = \mathbf{f}^{(m)} - \alpha \nabla J_I(\mathbf{f}^{(m)}) - \alpha \nabla J_S(\mathbf{f}^{(m)})$

# Distribution des données en radioastronomie

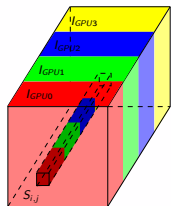
## Déconvolution

- Données : Images *dirty*  $\tilde{\mathbf{I}}_\ell$  du ciel à plusieurs longueurs d'onde  $\ell$
- Critère : attache aux données + **reg. spatiale** + **reg. spectrale**

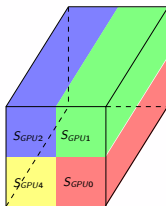
$$J(\mathbf{f}) = \sum_{\ell=1}^L \|\tilde{\mathbf{I}}_\ell - \mathbf{H}_\ell \mathbf{I}_\ell\|_2^2 + \mu_I \sum_{\ell=1}^L \|\mathbf{D}_I \mathbf{I}_\ell\|_2^2 + \mu_S \sum_{i,j=1,1}^{N^2} \|\mathbf{D}_S \mathbf{S}_{i,j}\|_2^2,$$

- Descente :  $\mathbf{f}^{(m+1)} = \mathbf{f}^{(m)} - \alpha \nabla J_I(\mathbf{f}^{(m)}) - \alpha \nabla J_S(\mathbf{f}^{(m)})$

## Distribution des données



par images  $\mathbf{I}_\ell$



par spectres  $\mathbf{S}_{i,j}$

# Distribution des données en radioastronomie

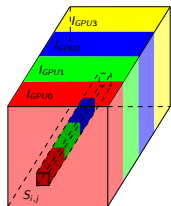
## Déconvolution

- Données : Images *dirty*  $\tilde{I}_\ell$  du ciel à plusieurs longueurs d'onde  $\ell$
- Critère : attache aux données + **reg. spatiale** + **reg. spectrale**

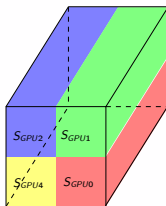
$$J(\mathbf{f}) = \sum_{\ell=1}^L \|\tilde{I}_\ell - \mathbf{H}_\ell \mathbf{I}_\ell\|_2^2 + \mu_I \sum_{\ell=1}^L \|\mathbf{D}_I \mathbf{I}_\ell\|_2^2 + \mu_S \sum_{i,j=1,1}^{N^2} \|\mathbf{D}_S \mathbf{S}_{i,j}\|_2^2,$$

- Descente :  $\mathbf{f}^{(m+1)} = \mathbf{f}^{(m)} - \alpha \nabla J_I(\mathbf{f}^{(m)}) - \alpha \nabla J_S(\mathbf{f}^{(m)})$

## Distribution des données



par images  $I_\ell$



par spectres  $S_{i,j}$

## Calcul de $\nabla J_S(\mathbf{f}^{(m)})$

- $N^2$  convolutions spectrales 1D

# Distribution des données en radioastronomie

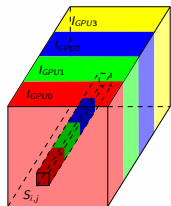
## Déconvolution

- Données : Images *dirty*  $\tilde{I}_\ell$  du ciel à plusieurs longueurs d'onde  $\ell$
- Critère : attache aux données + **reg. spatiale** + **reg. spectrale**

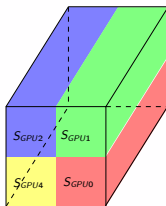
$$J(\mathbf{f}) = \sum_{\ell=1}^L \|\tilde{I}_\ell - \mathbf{H}_\ell \mathbf{I}_\ell\|_2^2 + \mu_I \sum_{\ell=1}^L \|\mathbf{D}_I \mathbf{I}_\ell\|_2^2 + \mu_S \sum_{i,j=1,1}^{N^2} \|\mathbf{D}_S \mathbf{S}_{i,j}\|_2^2,$$

- Descente :  $\mathbf{f}^{(m+1)} = \mathbf{f}^{(m)} - \alpha \nabla J_I(\mathbf{f}^{(m)}) - \alpha \nabla J_S(\mathbf{f}^{(m)})$

## Distribution des données



par images  $I_\ell$



par spectres  $S_{i,j}$

## Calcul de $\nabla J_S(\mathbf{f}^{(m)})$

- $N^2$  convolutions spectrales 1D
  - Stratégie de transferts mémoire
- #1 redistribution massive **avant et après**  
(CPU  $\leftrightarrow$  GPU ou interGPU)

# Distribution des données en radioastronomie

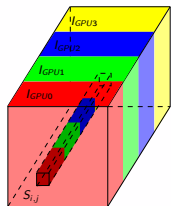
## Déconvolution

- Données : Images *dirty*  $\tilde{\mathbf{I}}_\ell$  du ciel à plusieurs longueurs d'onde  $\ell$
- Critère : attache aux données + **reg. spatiale** + **reg. spectrale**

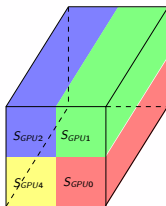
$$J(\mathbf{f}) = \sum_{\ell=1}^L \|\tilde{\mathbf{I}}_\ell - \mathbf{H}_\ell \mathbf{I}_\ell\|_2^2 + \mu_I \sum_{\ell=1}^L \|\mathbf{D}_I \mathbf{I}_\ell\|_2^2 + \mu_S \sum_{i,j=1,1}^{N^2} \|\mathbf{D}_S \mathbf{S}_{i,j}\|_2^2,$$

- Descente :  $\mathbf{f}^{(m+1)} = \mathbf{f}^{(m)} - \alpha \nabla J_I(\mathbf{f}^{(m)}) - \alpha \nabla J_S(\mathbf{f}^{(m)})$

## Distribution des données



par images  $\mathbf{I}_\ell$

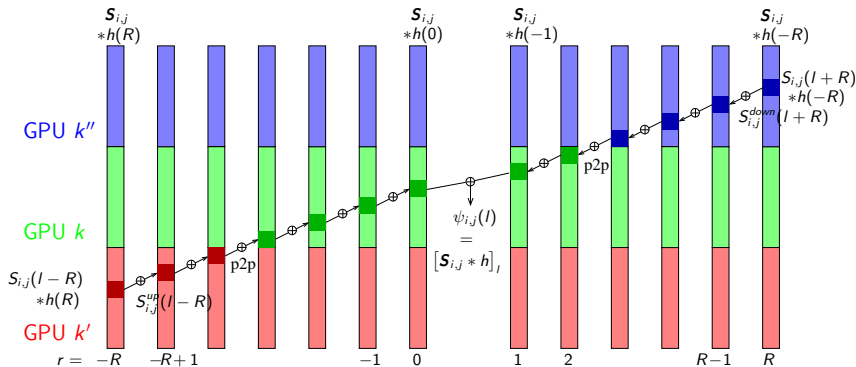


par spectres  $\mathbf{S}_{i,j}$

## Calcul de $\nabla J_S(\mathbf{f}^{(m)})$

- $N^2$  convolutions spectrales 1D
  - Stratégie de transferts mémoire
- #1 redistribution massive **avant et après** ( $\text{CPU} \leftrightarrow \text{GPU}$  ou *interGPU*)
  - #2 échanges interGPU de **résultats intermédiaires**

## Convolution spectrale distribuée (Olivier Pérard)



## Premiers résultats d'accélération

Xeon (28 coeurs)	$S_{zay}$ [1 GPU <sub>V100</sub> ]	$S_{zay}$ [8 GPU <sub>V100</sub> ]
119 s	11.0 s	1.58 s
<i>Matlab multithreads</i>	( $\times 10.8$ )	( $\times 7.0$ )

convolution d'un hypercube de taille  $1024^3$  pour un noyau de taille 512

- 1 Parcours depuis le doctorat
- 2 Adéquation Algorithme Architecture ( $A^3$ )
  - Problématique et positionnement
  - $A^3$  au sein du GPI
  - Encadrement de thèses et collaborations
- 3 Calcul intensif sur GPU/FPGA
  - Paire de projection/rétroprojection  $\mathcal{P}/\mathcal{R}$
  - Opérateur de convolution  $\mathcal{C}$
- 4 Calcul distribué sur serveur multiGPU
  - Reconstruction tomographique
  - DeconvSKA : Déconvolution en radioastronomie
- 5 Perspectives
  - Une méthodologie  $A^3$  à échelle humaine
  - SKA, un défi  $A^3$  à grande échelle



# Une méthodologie $A^3$ à échelle humaine

Faciliter le dialogue entre *algorithmiciens* et *architectes*

- vision synthétique **des algorithmes candidats**
- vision synthétique du **potentiel et des limitations des architectures**
- utilisation d'**outils simples** : analyse *off-line* des accès mémoire, *roofline model*...

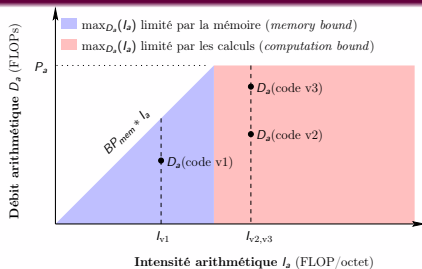
# Une méthodologie $A^3$ à échelle humaine

## Faciliter le dialogue entre *algorithmiciens* et *architectes*

- vision synthétique **des algorithmes candidats**
- vision synthétique du **potentiel et des limitations des architectures**
- utilisation d'**outils simples** : analyse *off-line* des accès mémoire, *roofline model*...

## *Roofline model*

- Utilisation plus systématique
- Extension pour FPGA, tensor cores...



# Terrain d'expérimentation de la démarche $A^3$

## Reconstruction tomographique

- Exploration algorithmique et architecturale des paires  $\mathcal{P}/\mathcal{R}$
- Fort potentiel de collaborations locales (ex : plateforme **TOMX**)

## Radioastronomie

- ROHSA-GPU - *séparation de source accélérée sur GPU*
- DeconvSKA - *déconvolution sur serveur multiGPU*

# Terrain d'expérimentation de la démarche $A^3$

## Reconstruction tomographique

- Exploration algorithmique et architecturale des paires  $\mathcal{P}/\mathcal{R}$
- Fort potentiel de collaborations locales (ex : plateforme **TOMX**)

## Radioastronomie

- ROHSA-GPU - *séparation de source accélérée sur GPU*
- DeconvSKA - *déconvolution sur serveur multiGPU*

## Utilisation des outils de l'IA

- CNN mimant algorithmes difficilement parallélisables

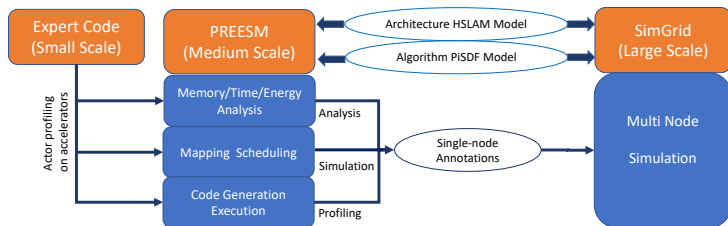
# Projet ANR DARK-ERA (1/2)

Le supercalculateur SDP (Science Data Processing), un projet HPC pionnier

- Flux de données gigantesque à traiter en temps réel
- Chaîne algorithmique complexe d'inversion de données
- Budget énergétique limité

Objectifs A<sup>3</sup> de DARK-ERA

#1 Outil de prototypage rapide SimSDP    #2 Exploration des accélérateurs *low power*



SimSDP construit à partir de PREESM et SimGrid

# Projet ANR DARK-ERA (2/2)

## Enjeux et motivations

### • Outil SimSDP

- Défi commun de son adoption par les radioastronomes
- Utilisation pour d'autres applications « flot de données »

### • Intérêt propre pour le GPI

- Prototype FPGA de la chaîne SDP de NenuFAR (Nançay)
- Complémentarité avec le projet ExaSKA

### • Interactions enrichissantes

Radioastronomie	↔	Problèmes inverses
Informatique HPC	↔	Electronique embarquée
A <sup>3</sup> « outils »	↔	A <sup>3</sup> « proche applications/architecture »

# Publications mentionnées dans la présentation

Liste complète : <https://l2s.centralesupelec.fr/u/gac-nicolas/publications/>

[R2019<sub>a</sub>] A. Marchal, M.A. Miville-Deschênes, F. Orioux, N. Gac, C. Soussen, M.J. Lesot, A. Revault d'Allonnes, Q. Salomé, ROHSA : Regularized Optimization for Hyper-Spectral Analysis, *Astronomy and Astrophysics - AA*, 2019

[R2019<sub>b</sub>] **C. Chapdelaine**, A. Mohammad-Djafari, N. Gac, E. Parra, Error-Splitting Forward Model for Iterative Reconstruction in X-ray Computed Tomography and application with Gauss-Markov-Potts prior, *IEEE Transactions on Computational Imaging*, 2019

[R2018<sub>a</sub>] **M. Martelli**, N. Gac, A. Merigot, C. Enderli, 3D Tomography back-projection parallelization on Intel FPGAs using OpenCL, *Journal of Signal Processing Systems*, Springer, 2018

[R2018<sub>b</sub>] **L. Wang**, A. Mohammad-Djafari, N. Gac, **M. Dumitru**, 3D X-ray Computed Tomography with a Hierarchical Prior model for Sparsity in Haar Transform domain, *Entropy, Special Issue Probabilistic Methods for Inverse Problems*, MDPI, 2018

[R2017<sub>a</sub>] **L. Wang**, A. Mohammad-Djafari, N. Gac, X-ray Computed Tomography using a sparsity enforcing prior model based on Haar transformation in a Bayesian framework, *Special Issue of Fundamenta Informaticae*, IOS Press, 2017

[R2017<sub>b</sub>] **C. Chapdelaine**, A. Mohammad-Djafari, N. Gac, E. Parra, A 3D Bayesian Computed Tomography Reconstruction Algorithm with Gauss-Markov-Potts Prior Model and its Application on Real Data, *Special Issue of Fundamenta Informaticae*, IOS Press, 2017

[R2014<sub>a</sub>] **N. Chu**, J. Picheral, A. Mohammad-Djafari, N. Gac, A robust super-resolution approach with sparsity constraint in acoustic imaging, *Applied Acoustics*, Elsevier, 2014, 76, pp.197-208.

[R2014<sub>b</sub>] F. Schmidt, I. Shatalina, M. Kowalski, N. Gac, B. Saggin, et al., Toward a numerical deshaker for PFS, *Planetary and Space Science*, Elsevier, 2014, 91, pp.45 - 51.

[R2013] **T. Boulay**, N. Gac, A. Mohammad-Djafari, J. Lagoutte, Algorithmes de reconnaissance NCTR et parallélisation sur GPU, *Traitement du Signal*, Lavoisier, 2013, 6, pp.309-342.

[R2012] M.L. Gallin-Martel., Y. Grondin, N. Gac et al, Experimental results and first <sup>22</sup>Na source image reconstruction by two prototype modules in coincidence of a liquid Xenon Positron Emission Tomograph for small animal imaging, *Nuclear Instruments and Methods in Physics Research*, Elsevier, 2012, 682, pp.66-74

[R2009] N. Gac, S.Mancini, M. Desvignes et D. Houzet, High Speed 3D Tomography on CPU, GPU and FPGA, *EURASIP Journal on Embedded systems*, SpringerOpen, 2009

# Publications mentionnées dans la présentation

Liste complète : <https://l2s.centralesupelec.fr/u/gac-nicolas/publications/>

- [C<sub>2020a</sub>] **D. Diakite**, **M. Martelli**, **N. Gac**, An OpenCL pipeline implementation on Intel FPGA for 3D backprojection, *International Conference on Image Formation in X-Ray Computed Tomography, 2020, Regensburg*
- [C<sub>2020b</sub>] **M. Seznec**, **N. Gac**, F. Orioux, A. Sashala Naik, An efficiency-driven approach for real-time optical flow processing on parallel hardware, *IEEE International Conference on Image Processing (ICIP), 2020, Abu Dhabi*
- [C<sub>2020c</sub>] **M. Chghaf**, **N. Gac**, Data distribution on a multi-GPU node for TomoBayes CT reconstruction, *IEEE Conference on Embedded and Real-Time Computing Systems and Applications (RTSCA), 2020, South Korea*
- [C<sub>2020d</sub>] **M. Seznec**, **N. Gac**, F. Orioux, A. Sashala Naik, A new convolutions algorithm to leverage tensor cores, *GPU Technology Conference (GTC), May 2020, Silicon Valley, United States*
- [C<sub>2019a</sub>] **N. Georgin**, **C. Chapdelaine**, **N. Gac**, A. Mohammad-Djafari, E. Parra, Multi-streaming and multi-GPU optimization for a matched pair of Projector and Backprojector, *2019 International Conference on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine, Jun 2019, Philadelphia, United States*
- [C<sub>2018a</sub>] **C. Chapdelaine**, **N. Gac**, A. Mohammad-Djafari, E. Parra, New GPU implementation of Separable Footprint (SF) Projector and Backprojector : first results, *International Meeting on Image Formation in X-Ray Computed Tomography, Salt Lake City, US, 2018*
- [C<sub>2018b</sub>] **M. Seznec**, **N. Gac**, A. Ferrari, F. Orioux, A Study on Convolution Using Half-Precision Floating-Point Numbers on GPU for Radioastronomy Deconvolution, *IEEE SIPS, Cape Town, South Africa, October 2018*
- [C<sub>2017d</sub>] **M. Dumitru**, **N. Gac**, **L. Wang**, A. Mohammad-Djafari, Unsupervised sparsity enforcing iterative algorithms for 3D image reconstruction in X-ray Computed Tomography, *Fully3D, 2017*
- [C<sub>2014b</sub>] **N. Chu**, **N. Gac**, J. Picheral, A. Mohammad-Djafari, 2D Convolution model using (in)variant kernels for fast acoustic imaging, *BEDEC 2014, Berlin Beamforming Conference, 15 p., 2014*
- [C<sub>2014d</sub>] **L. Chen**, T. Rodet, **N. Gac**, A simple and efficient super-short-scan algorithm of fan-beam reconstruction for multiple circular trajectories : solution towards the truncated data, *CT Meeting, Salt Lake City, pp. 212-215, 2014*
- [C<sub>2013d</sub>] **L. Chen**, T. Rodet, **N. Gac**, A penalized weighted least-squares image reconstruction based on scatter correction methods for X-ray CT, *2013 IEEE NSS and MIC, Seoul, 2013*
- [C<sub>2011a</sub>] **N. Gac**, A. Vabre, A. Mohammad-Djafari, Multi GPU parallelization of 3D bayesian CT algorithm and its application on real foam reconstruction with incomplete data set, *FVR, Poitiers, 2011*



Merci de votre attention