

Astroparticle Data Analysis  
Modeling, Interpretation, Simulation

**ADAMIS**

Analyse de Données en Astroparticules,  
Modélisation, Interprétation, Simulation

# Plan

- **Concept, équipe et mode de fonctionnement**
- **Exemples de réalisations et projets**
- **Prospective**

# Sophistication des données et des modèles

Rares

(peu de données)

Indépendantes

(un seul jeu de données pour une observable)

Privées

(pas d'accessibilité aux données externes)

Simple

(interprétation plutôt que traitement)



Nombreuses

traitement statistique

Interdépendantes

traitement joint

Publiques

analyse découplée de l'expérimentation

Complexes

méthodes innovantes

# Traitement de données à l'APC

Objectif: une équipe performante pour le traitement et l'analyse des données

- Maîtrisant les méthodes de traitement modernes (récentes!)
- Compétitive pour analyser les données en libre accès
- Capable de publier des données de qualité (!)
- Capable de développer des solutions innovantes à des problèmes complexes
- Capable de mettre en œuvre ces solutions pour leur application à des problèmes concrets

# Points clés

## Expériences

### Accès aux données

- Participation aux expériences
- Utilisation centres de données
- Collaborations

## Physique

### Modélisation

- Maîtrise des problématiques
- Modèle de la mesure  
 $d = f(\text{instrument, phénomène...})$
- Simulation numérique

### Solution formelle

- Mathématiques appliquées
- Statistique
- Traitement de l'information
- Traitement de signal et d'image

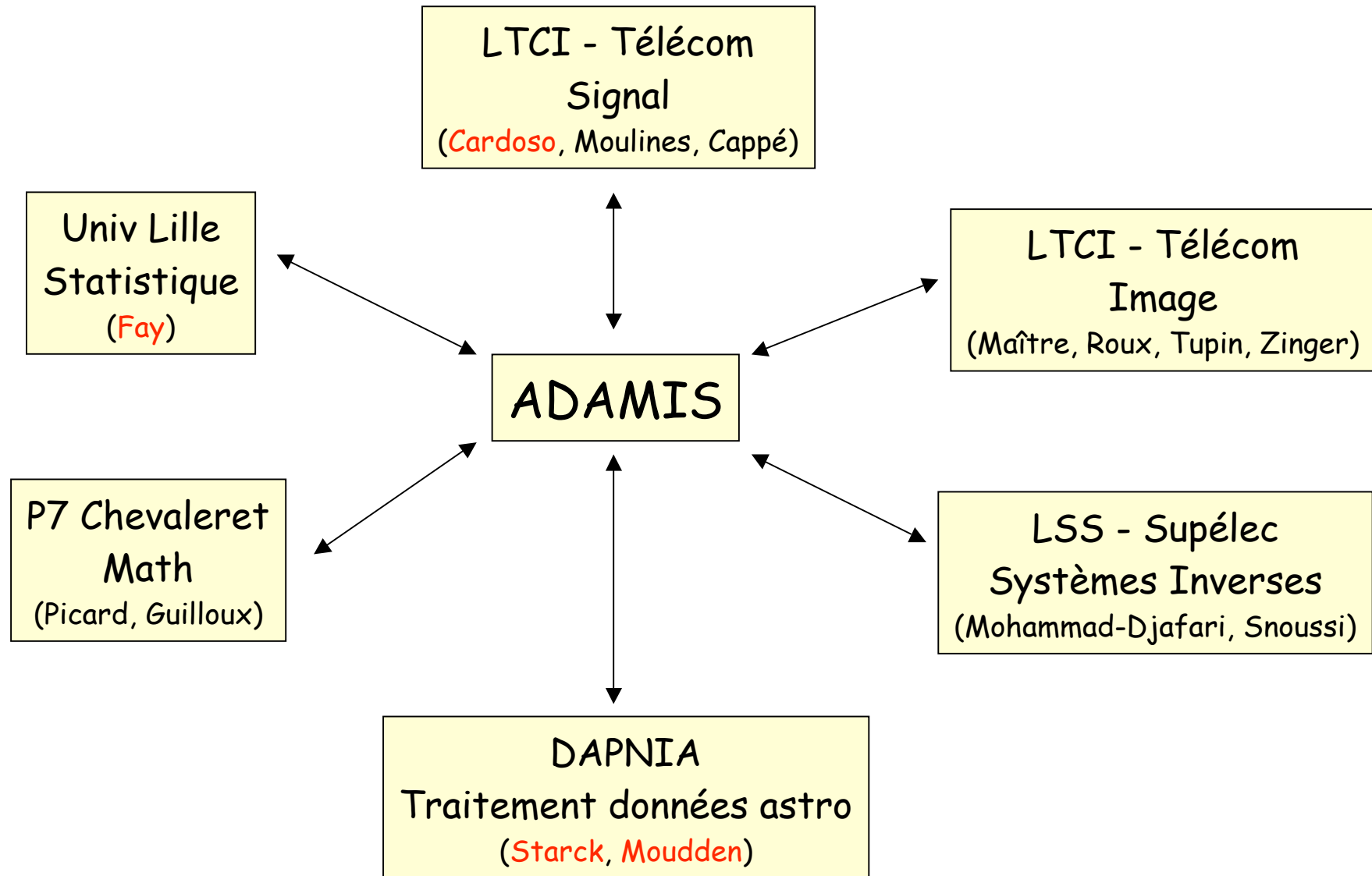
### Implémentation

- Algorithmique/Logiciel
- Bibliothèque de logiciels
- Moyens de calcul
- Production

## Mathématiques

## Informatique

# Pluridisciplinarité / Interfaces



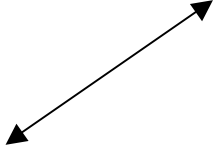
# Connections

Observatoire Virtuel

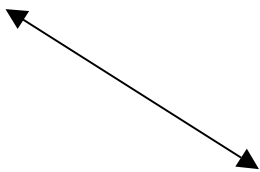


ADAMIS

Simulations  
Numériques



data GRID  
GRID computing



# Mode de fonctionnement

- 1- Identification de problèmes complexes (et intéressants sur le plan scientifique !) en terme de traitement/analyse de données
- 2- Identification des sous-taches et des verrous méthodologiques
- 3- Recherche de partenaires (communautés STIC, maths, physique), définition de projets et d'une équipe (souple)
- 4- On bosse... On publie... On applique... On re-publie...



# L'équipe ADAMIS aujourd'hui

## Membres APC

- 2003 ← **Éric Aubourg** (Ingénieur-chercheur CEA)  
2003 ← **Tristan Beau** (MC Paris 7)  
          **Jacques Delabrouille** (CR CNRS section 03)  
          **Éric Lesquoy** (Ingénieur-chercheur CEA)  
2003 ← **Fabrice Dodu** (IR Paris 7)

## Membres Associés

- Jean-François Cardoso** (CNRS ENST Paris)  
**Gilles Fay** (MC Lille Laboratoire de Statistique)  
**Jean-Luc Starck** (Ingénieur-chercheur CEA)

## Postdocs et Étudiants

- Patricio Vielva** (Postdoc PCC)  
**Yassir Moudden** (Postdoc CEA)

## Projets (achevés - en cours - idées...)

- Estimation spectrale multicomposantes
- Rééchantillonnage et échantillonnage inhomogène sur la sphère
- Séparation simultanée de composantes diffuses et ponctuelles
- Estimation spectrale et statistique sur la sphère
- Base de données avec interface internet multi-utilisateurs
- Application et développement de méthodes d'apprentissage
- Fusion de données CMB
- Emissions astrophysiques diffuses et observatoire virtuel
- Représentation de PSF et bruits inhomogènes
- Déconvolution et hyper-résolution en présence de bruit
- ...

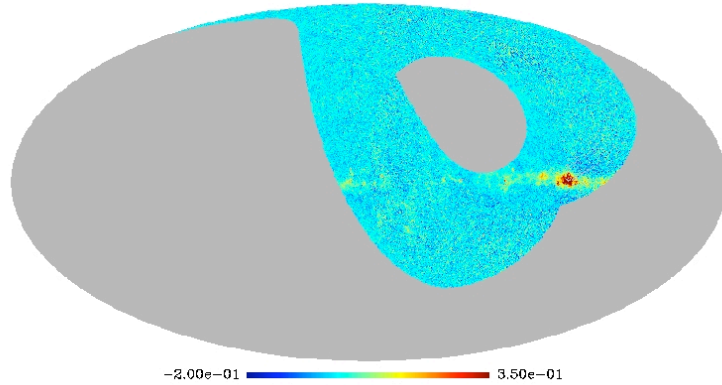
# Plan

- Concept, équipe et mode de fonctionnement
- Exemples de réalisations et projets
- Prospective

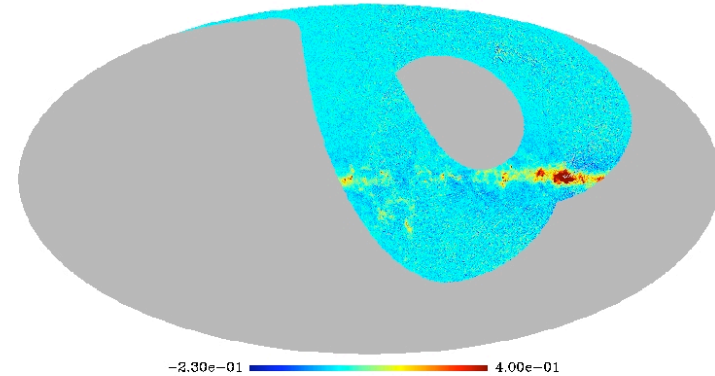
Un exemple de réalisation

# Observations Archeops

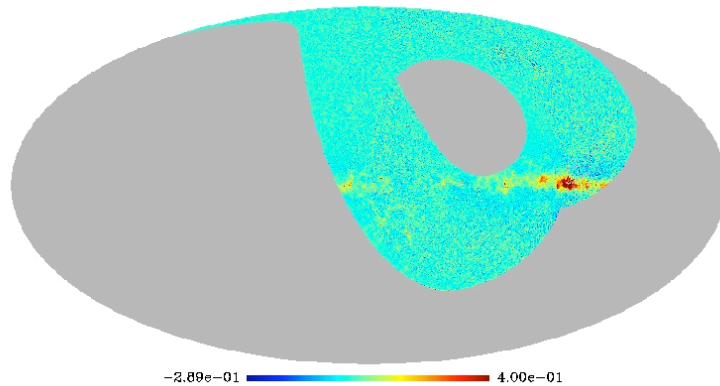
143k03



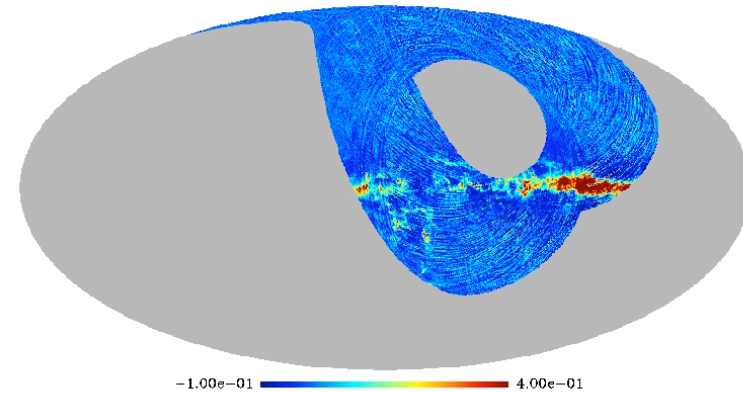
217k06



353k02



545k01



14 cartes au total dans 4 canaux: 3x143, 5x217, 5x353, 1x545

## Modèle

- chaque détecteur voit un mélange linéaire de plusieurs composantes
- On ne suppose rien sur les composantes sauf leur indépendance
- On cherche le nombre de composantes, leur spectre spatial de puissance
- On cherche aussi à déterminer la matrice de mélange
- On ne suppose rien sur les bruits des différents capteurs

$$Y_d(\theta, \varphi) = A_{dc} S_c(\theta, \varphi) + N_d(\theta, \varphi)$$

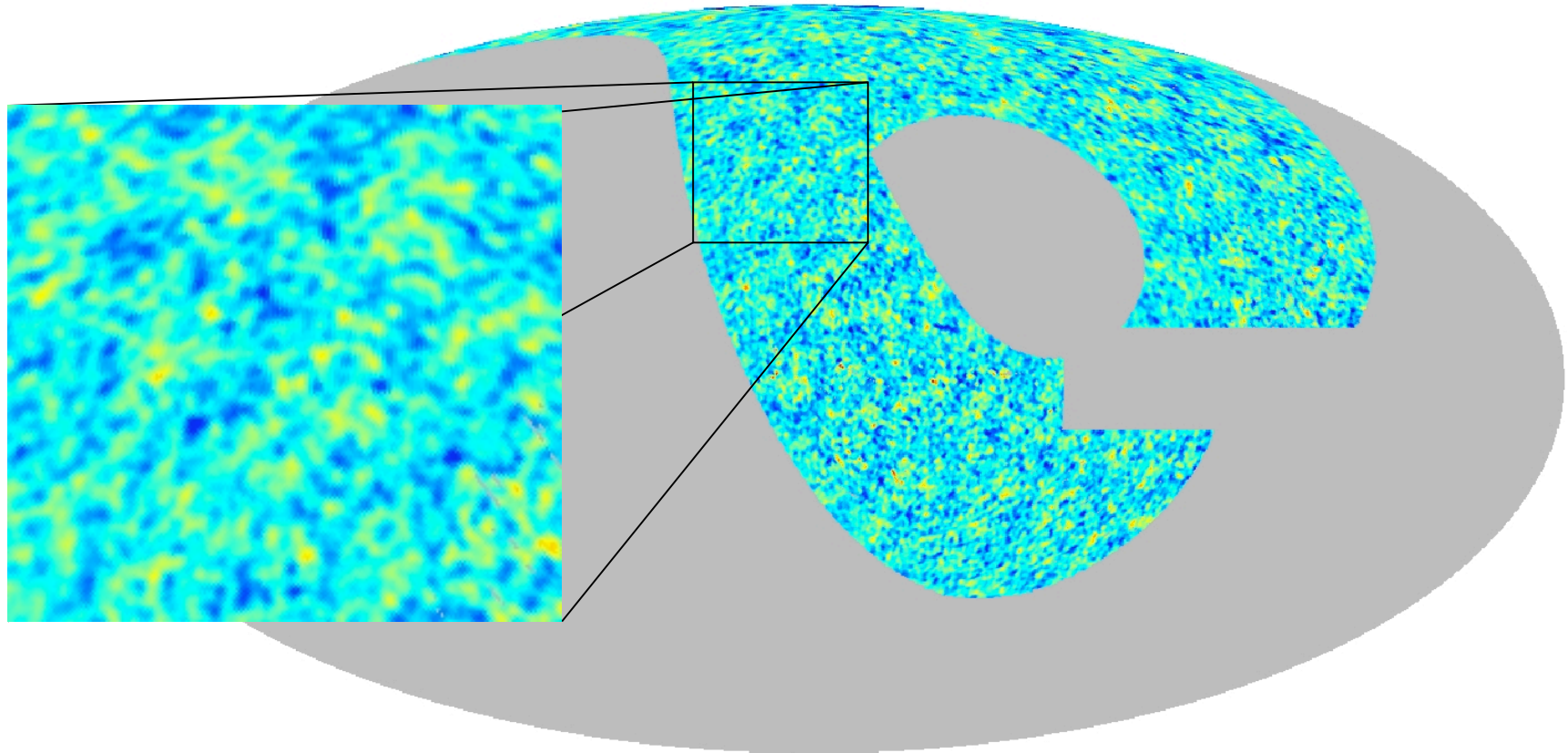
## Méthode (SMICA)

- ajustement des paramètres du modèle ( $A$ ,  $C_i$ (composantes),  $C_i$ (bruits)) par maximisation d'une vraisemblance approchée
- séparation de composante par Wiener utilisant les paramètres estimés

## Implantation

- Maximisation de la vraisemblance par algorithme EM suivi (pour accélérer la convergence) par une méthode de descente BFGS, correction de réponse par Monte-Carlo (pas très satisfaisant...), filtrage de Wiener

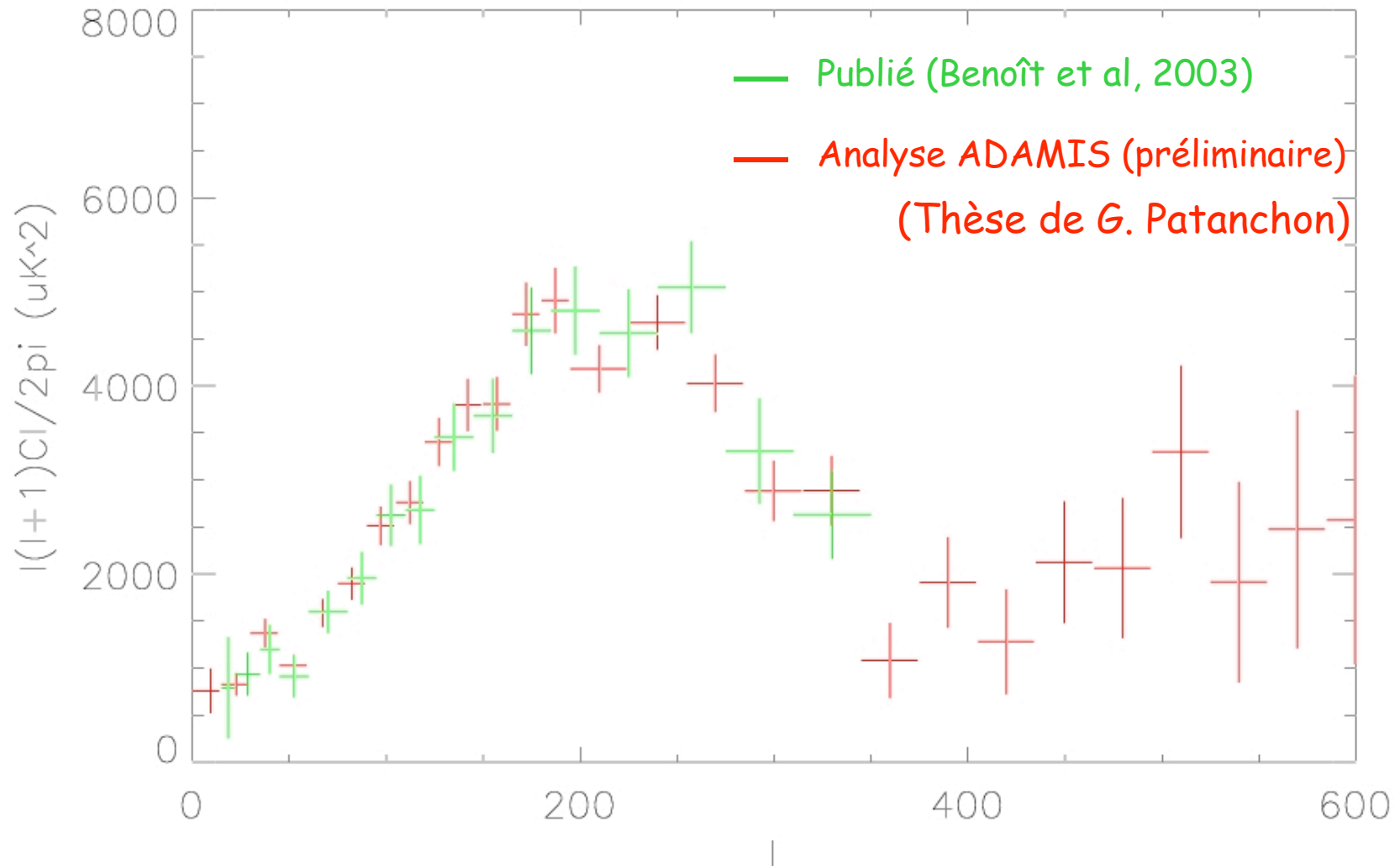
# Première composante : CMB



$-9.36e-05$    $1.52e-04$

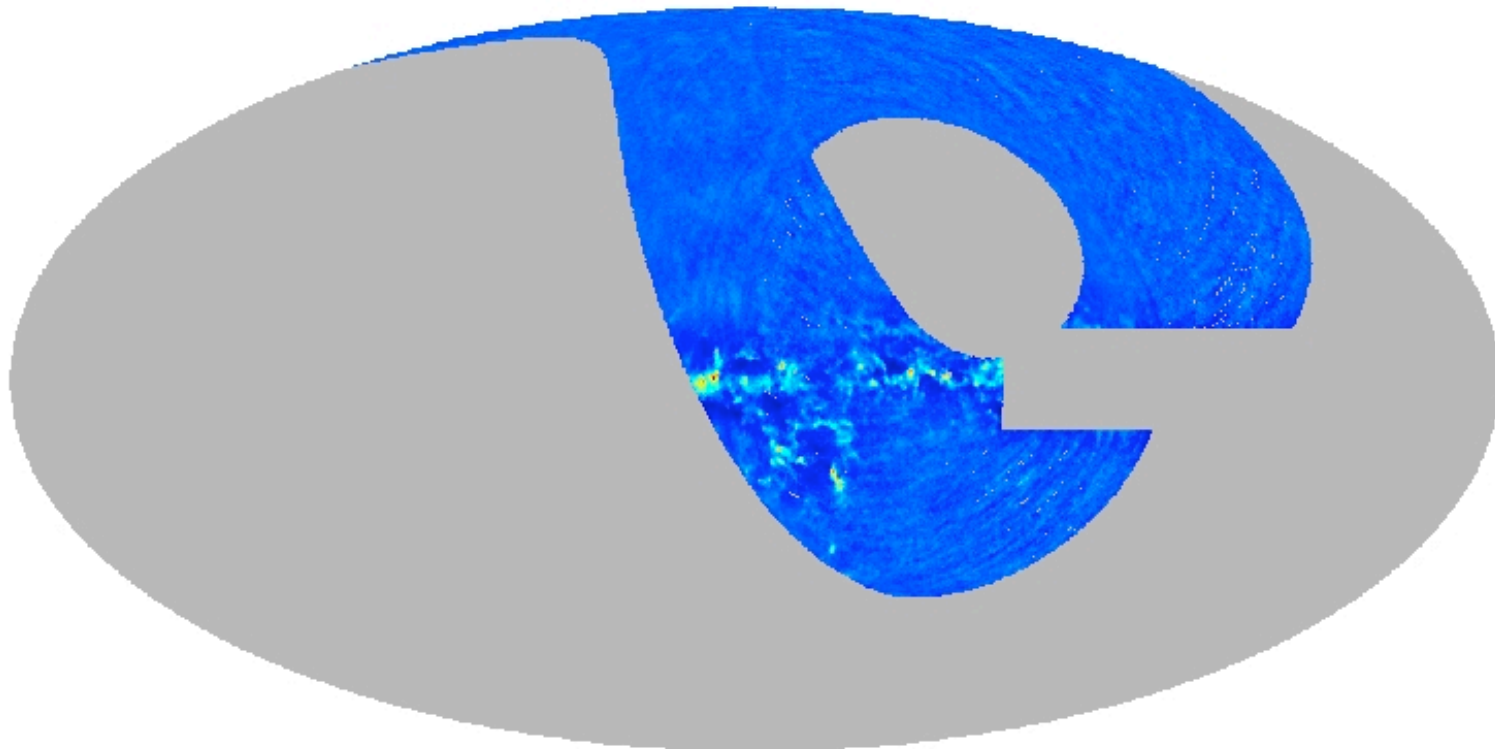
$T_{CMB}$

# Spectre de puissance du CMB





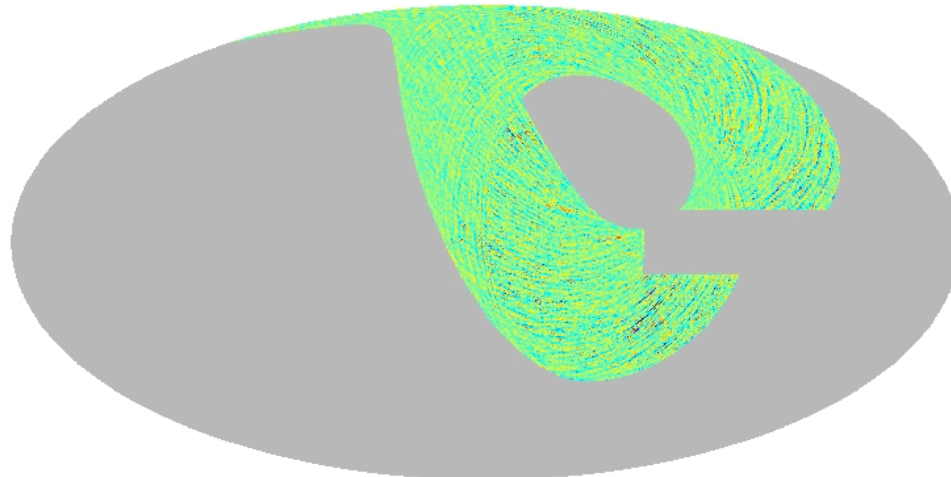
# Seconde composante : Galaxie



-1.01e-04  4.31e-04

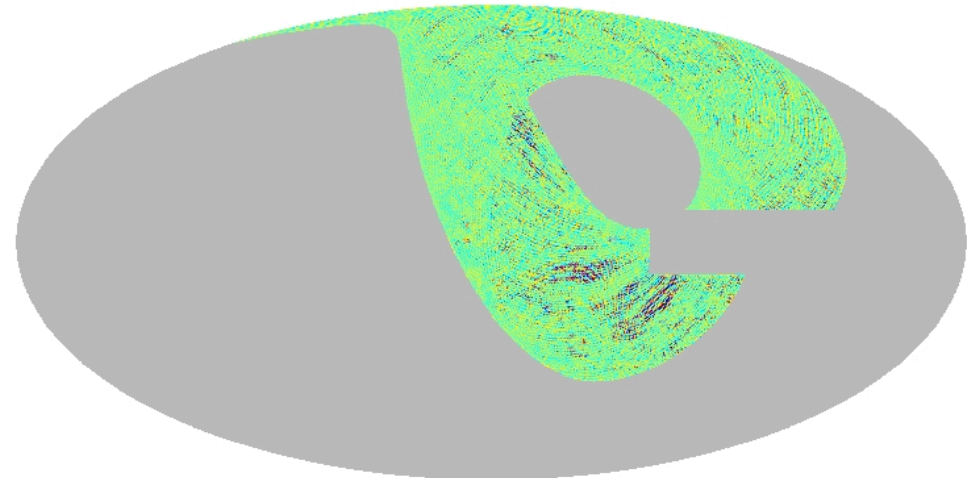
# Composantes « poubelle »

Composante ozone 1



-6.00e-01  6.00e-01

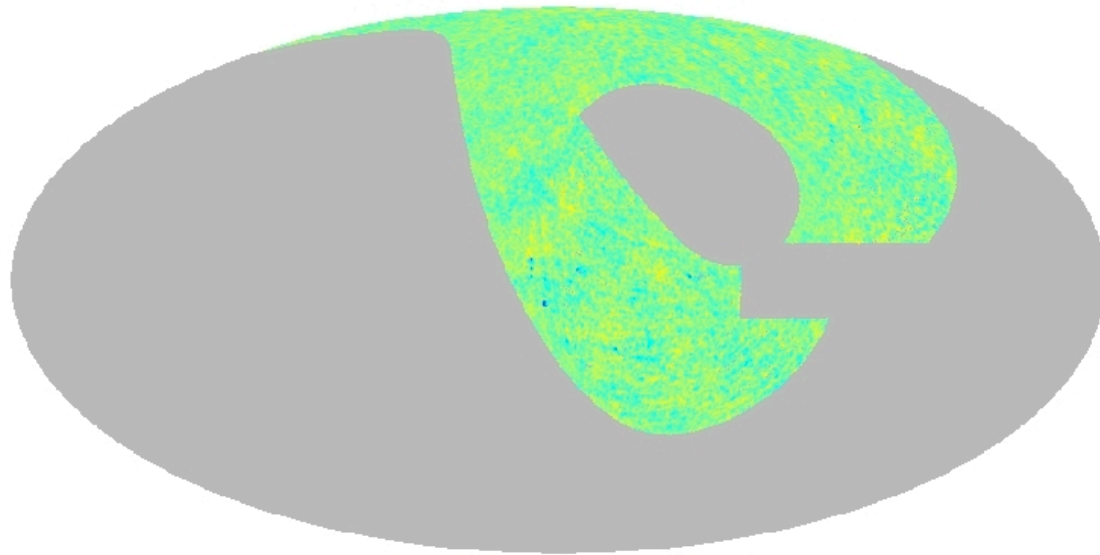
- Stries résiduelles
- Moins marquées dans les régions bien vues
- Probablement des résidus d'atmosphère



-4.00e-01  4.00e-01 mK, 545 GHz

# Comparaisons

WMAP-Archeops



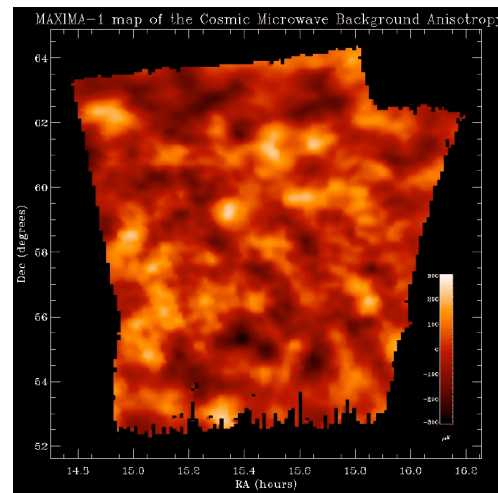
-6.00e-01  6.00e-01 mK CMB

WMAP et Archeops

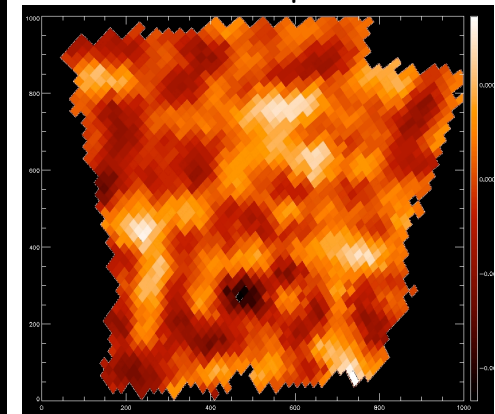
Somme et différence

Comparaison  
Archeops et MAXIMA

MAXIMA



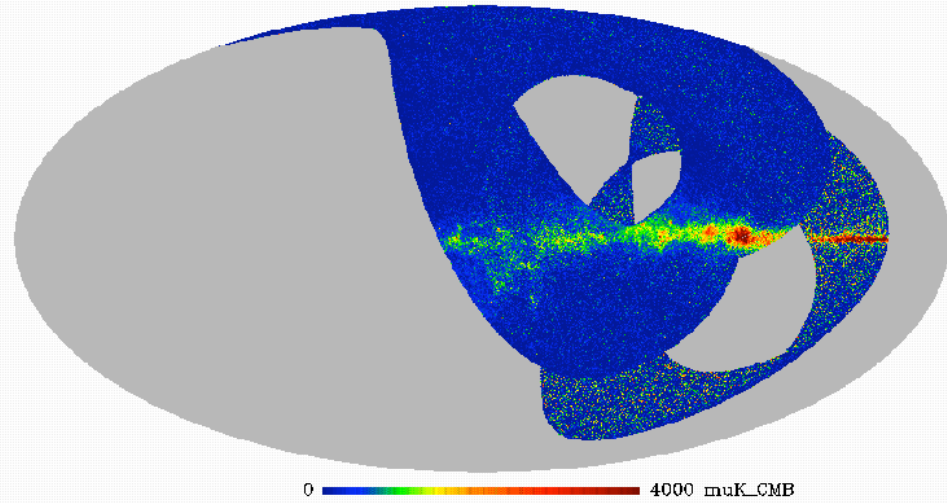
Archeops



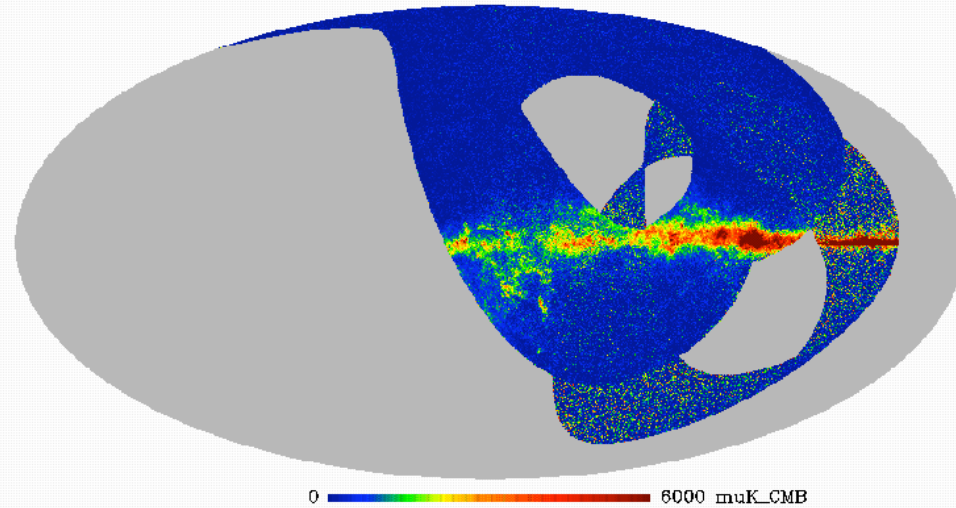
Suite naturelle  
fusion d'observations diffuses

# Fusion de cartes Archeops seul...

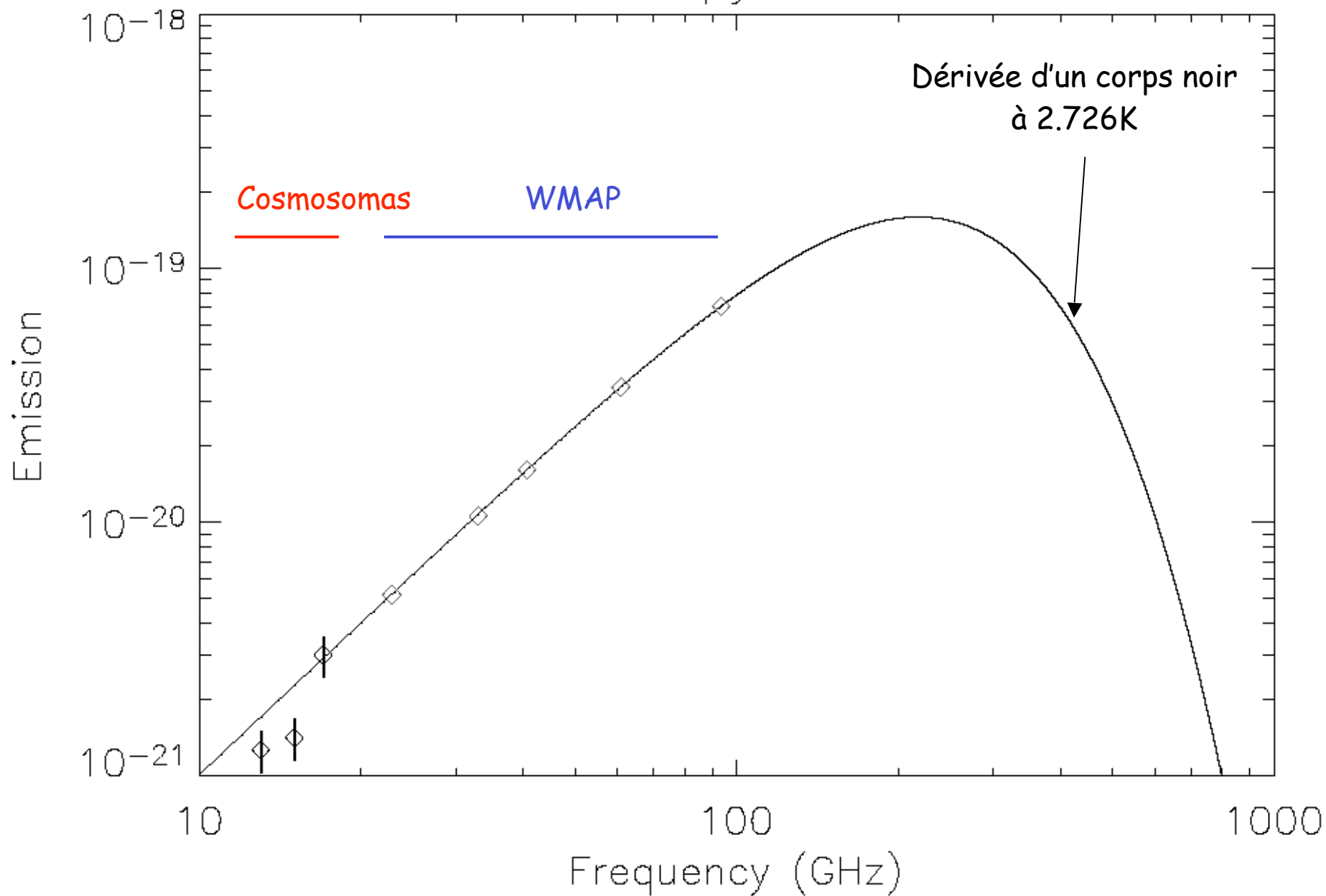
ARCHEOPS TRAPANI+KS1+KS3 FLIGHTS, GALAXY MAP @ 143 GHz



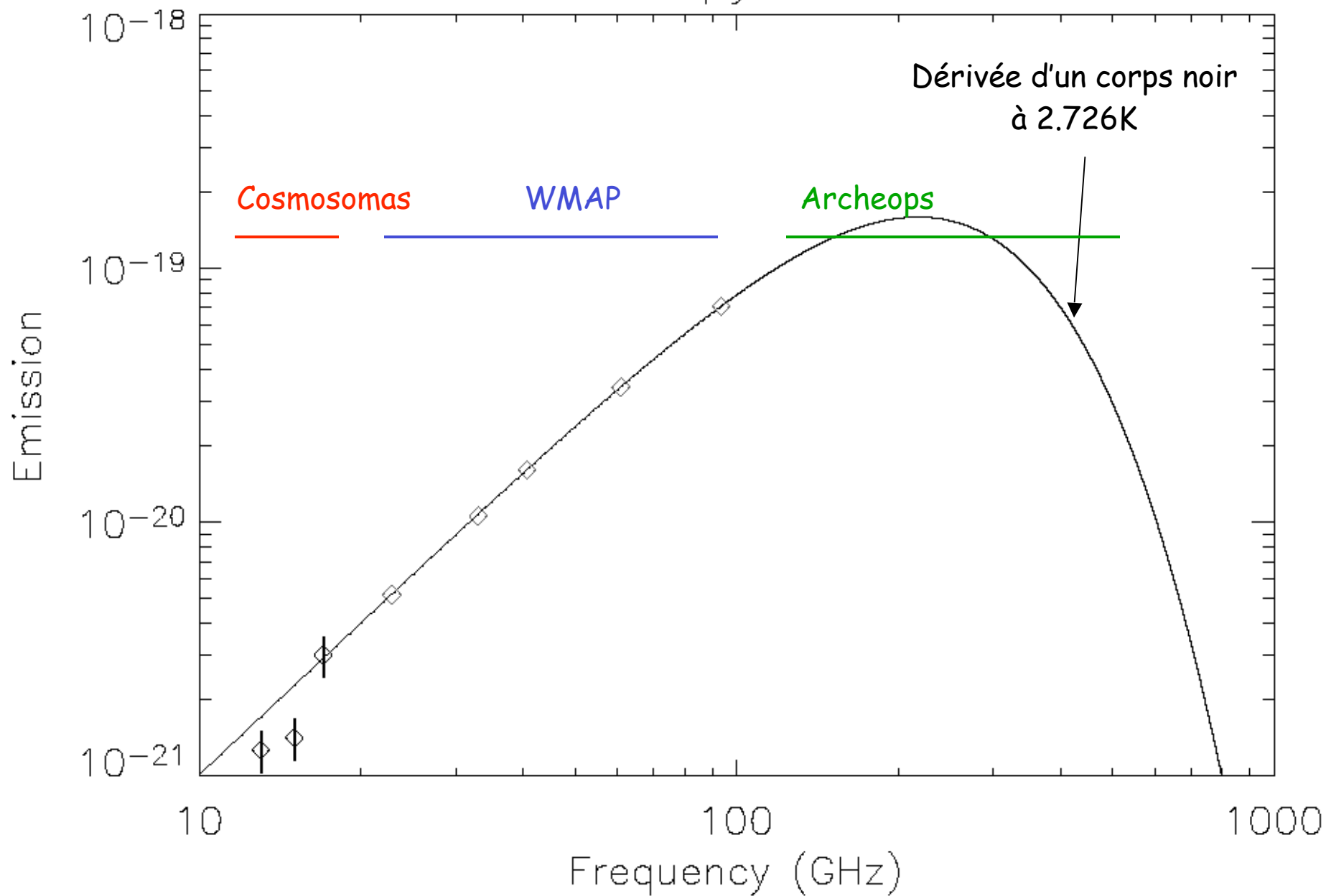
ARCHEOPS TRAPANI+KS1+KS3 FLIGHTS, GALAXY MAP @ 217 GHz



# CMB anisotropy emission law



# CMB anisotropy emission law



## Fusion de cartes multi-fréquence multi-résolution

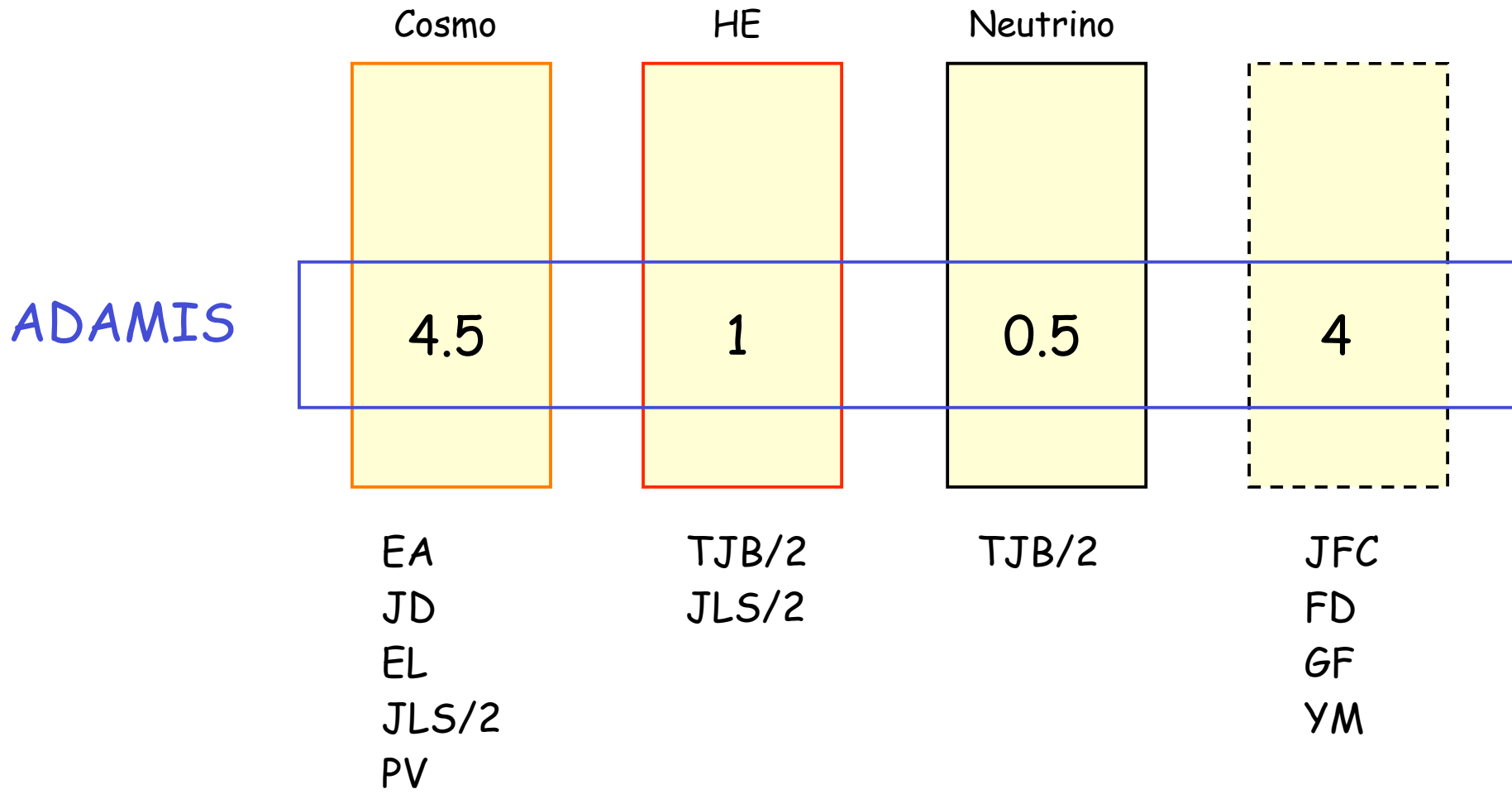
- Archeops + WMAP + Cosmosomas + MAXIMA + FIRAS + IRAS + DIRBE + Boomerang + VSA...
- Production de cartes  $\Delta T(\theta, \varphi, \lambda)$ , erreurs, corrélations des erreurs, source majeure de l'erreur, poids des données, noyau de convolution
- Production de cartes de CMB, poussière, SZ, synchrotron, résidus...



# Plan

- Concept, équipe et mode de fonctionnement
- Exemples de réalisations et projets
- **Prospective**

# Transversalité ?



Majeure / Mineure ?

# Développer la transversalité

## Cosmo: activité initiée il y a 3 ans!

- DPC et analyse Planck (avec la collaboration Planck)
- Retraitement des données Archeops (avec la collaboration Archeops)
- Fusion de données (avec les collaborations Cosmosomas, VSA, données WMAP...)
- Amas SZ (avec le groupe Cosmo)

## Pistes de développement:

- Analyse AUGER
  - Base de données (T. Beau, E. Aubourg, E. Lesquoy)
  - Statistique/recherche de sources (T. Beau + AUGER)
  - Classification/ajustement par apprentissage (avec maths)
- Detection ondes gravitationnelles (avec A. Buonanno + traitement signal + maths)
- EUSO
  - Analyse multi-longueur d'onde radio-IR-X-gamma (avec R. Terrier)
  - Calcul: évaluation G5 (E. Lesquoy, E. Aubourg avec LAL et DAPNIA, CC?)
  - Simulation et Monte-Carlo (avec M. Tagger, groupes APC, IAP, SAp, observatoire)

# Transfert de connaissance

- Promouvoir la qualité pour le traitement de données
  - Discussions - Transversalité
  - École d 'été
  - Cours (e.g. T. Beau cours C++ au PCC)
  - Interface (physique) - (math, signal, image, statistique, informatique)
  - Application pratiques dans les expériences
- Enseignement à Paris 7...
  - T. Beau
  - Modules d'enseignement?

# Equilibre ?

## Accès aux données

- Expériences OK
- Collaborations OK
- Gestion de données faible
- Recruter informaticien BDD

## Modélisation

- Plutôt fort surtout en Cosmo
- Développer hors cosmo
- Simulation faible à développer

## Solution formelle

- Plutôt fort
- Beaucoup de collaborateurs
- Bonnes perspectives

## Implémentation

- Faible
- 1 IR informatique insuffisant
- Moyens de calcul insuffisants
- Application: étudiants postdocs

## Problème (assez classique)

Déséquilibre entre idées/projets, et capacité de les réaliser;  
Déséquilibre entre développements théoriques et application;

- Besoin de postdocs et d'étudiants (développements/applications)
- Besoin d'informaticiens (applications/fonctionnement)
- Besoin de moyens informatiques et de leur administration

## Solutions (tout aussi classiques)

- Recruter et mettre les moyens en place
- Rester modeste, limiter les activités, définir les priorités ...

# Difficultés / Besoins

## 1- Besoin d'ingénieurs informaticiens

- Implantation des chaînes de traitement (exemple: prise en charge de la totalité ou d'une partie du traitement des données d'une expérience spatiale)
- Simulation de données (interface simulation numérique)
- Portabilité et maintenance de bibliothèques logiciel
- Optimisation numérique et calcul massivement parallèle (interface GRID)
- Implantation de bases de données
- Participation à l'Observatoire Virtuel français (interface OV)

## Difficultés / Besoins (suite)

### 2- Besoin de moyens de calcul semi-lourds et lourds

1. Moyens de proximité semi-lourds (exemple, 32-64 processeurs rapides, mémoire partagée 64-128 GO, disque RAID 20 TO)
2. Moyens de calcul lourds partagés (exemple 1024 processeurs, mémoire 500 GO, disque 100 TO, stockage 400 TO) pour le traitement de données et la simulation numérique en astroparticule et cosmologie



# Difficultés / Besoins (fin)

## 3- Capacité d'accueil

Accueil de spécialistes de jeux de données ou de chercheurs d'autres disciplines : postdocs et visiteurs

Infrastructures d'accueil

Postdocs (1-2/an)  
Etudiants (message aux écoles doctorales)  
Ingénieurs info (3 IR *maintenant ou bientôt !*)  
Moyens informatiques propres (200 k€/4ans + 15 k€/an)  
Moyens informatiques partagés (?)  
Chercheurs (de temps en temps)  
Soutien des tutelles (mobilité, concentration)

# Risques ?

- **Plus complexe que nécessaire? Projets bien/mal ciblés? Innovant?**
  - Publications**
  - Collaborations**
  - Applications**
  - Evaluation à 2 ans (2005), à 4 ans (2007)**
  
- **Structure qui ne fonctionne pas?**
  - Souplesse !**
  - Les investissements sont durables même dans une autre structure**