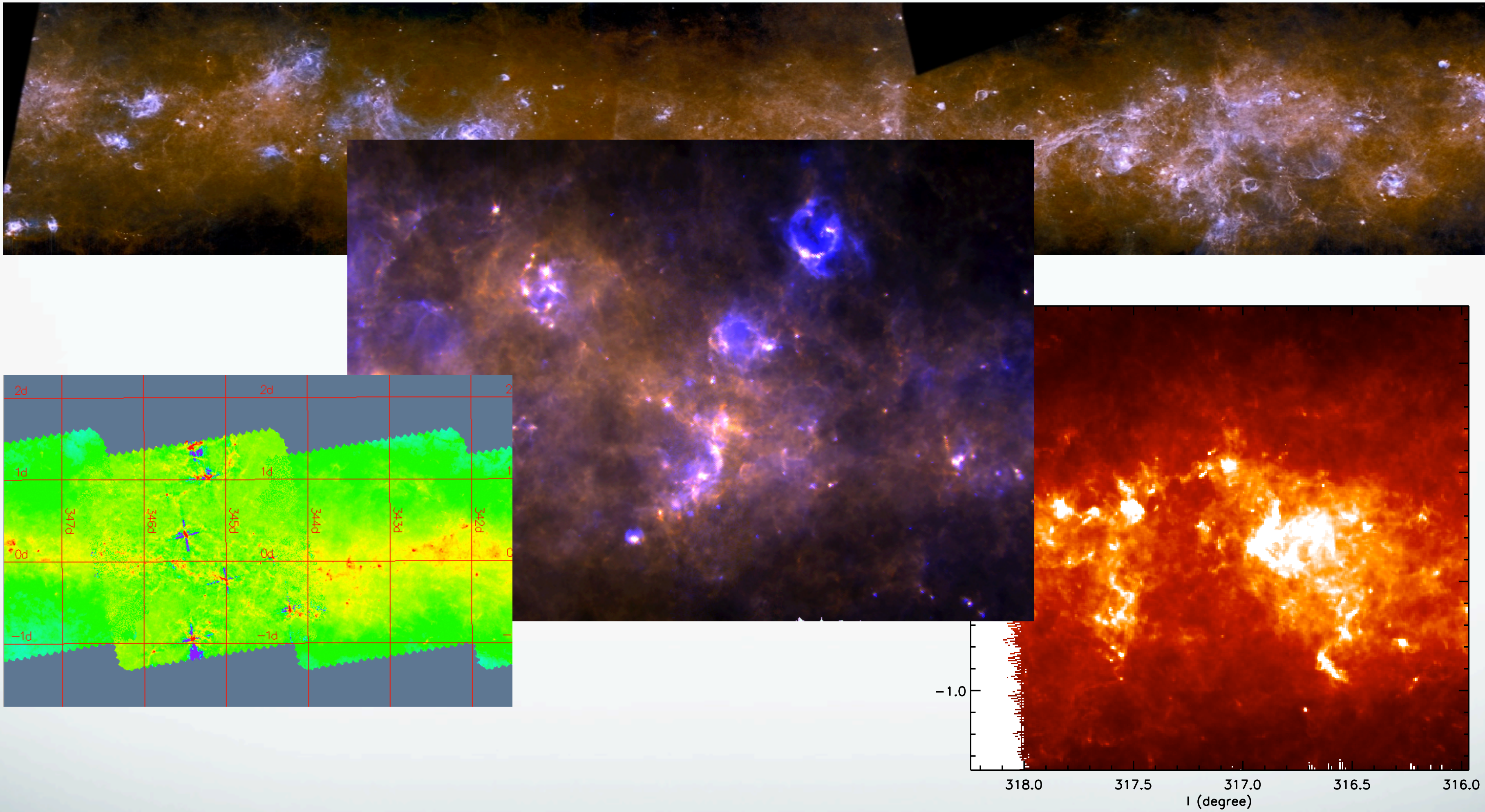
ROMAGAL: the Hi-GAL data reduction pipeline



A. Traficante and the Hi-GAL map-making team

MANCHESTER
1824

Herschel map-making workshop 28-31/01/13

Outline

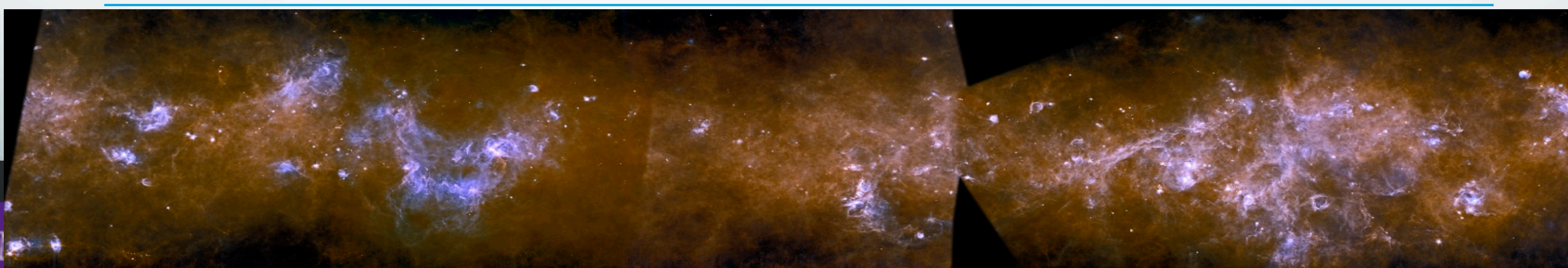
- Hi-GAL survey
- Hi-GAL pipeline: pre-processing steps
- Hi-GAL pipeline: ROMAGAL algorithm
- ROMAGAL PGLS
- Map calibration (Zero-level offset)
- Conclusions

Hi-GAL survey

Hi-GAL <u>Herschel Infrared Galactic Plane Survey</u> (Molinari et al. 2010)

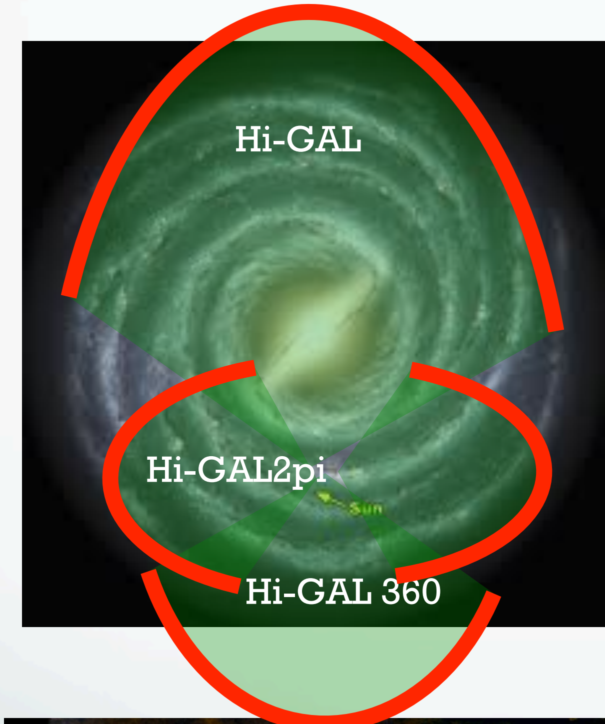
The largest Herschel open-time key project

Wavelength Coverage	70-500 μm
Sky coverage	2° x 360° |b| ≤ 1° (following the Galactic warp)
Total Herschel time	~ 900 hours
Scanning strategy	Parallel (PACS and SPIRE observe simultaneously)
Scan speed	60''/sec



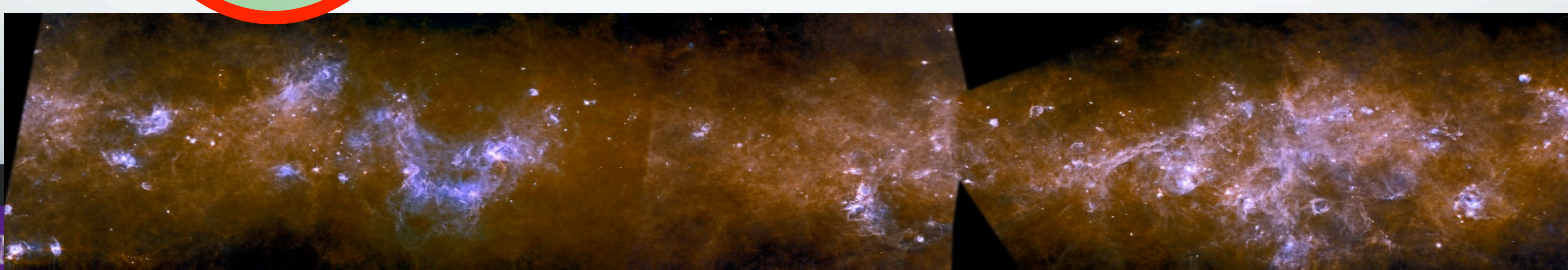
Hi-GAL survey

Hi-GAL <u>Herschel Infrared Galactic Plane Survey</u> (Molinari et al. 2010)



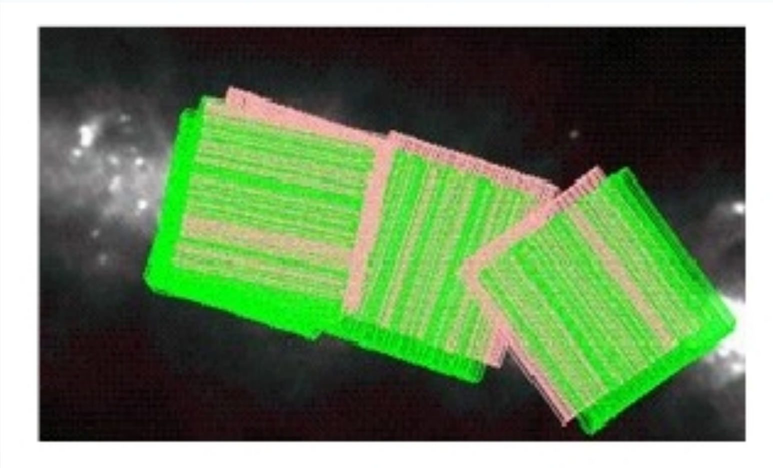
Hi-GAL 1: $|b| < 1°$ $-71° < l < 66°$

Hi-GAL 360 & HiGAL2pi: the whole Plane

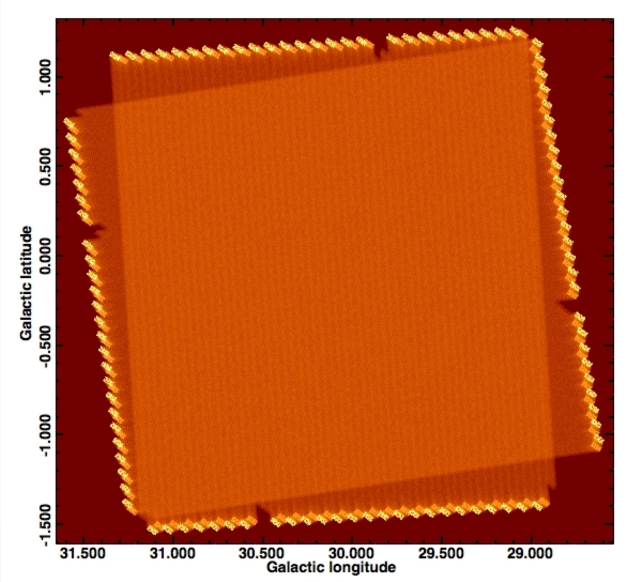


Hi-GAL survey

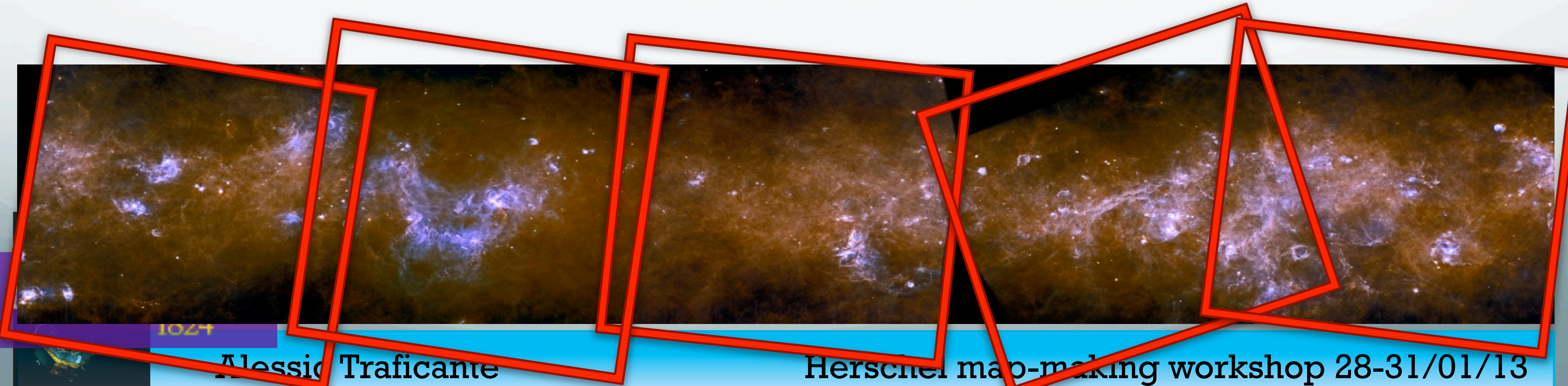
Hi-GAL <u>Herschel Infrared Galactic Plane Survey</u> (Molinari et al. 2010)



Tiles cover the Galactic Plane
with different orientations



PACS 70 μm l=30 coverage



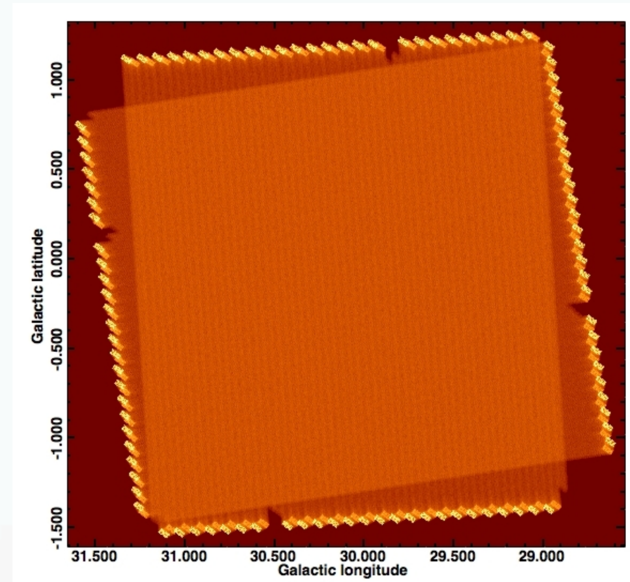
Hi-GAL survey

Hi-GAL <u>Herschel Infrared Galactic Plane Survey</u> (Molinari et al. 2010)

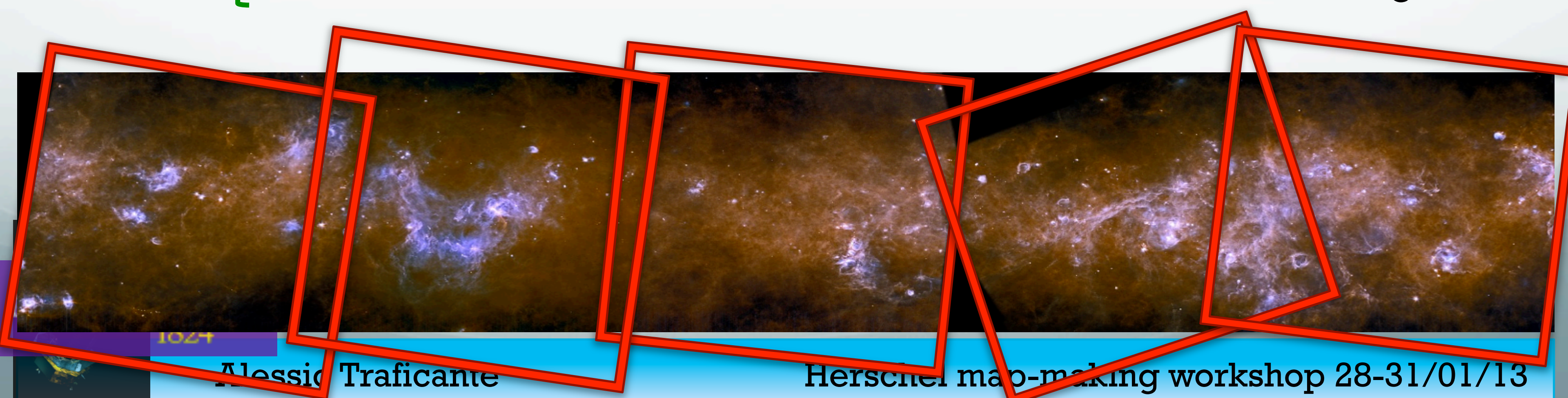
Each tile:

- 2° x 2°
- ~ 54 scan-legs (nomimal+orthogonal)
- ~ 10-30 samples per pixel (depending from the map resolution)

PIXELS
- ~ 3000 x 3000 (PACS 70 μ m)
- ~ 500 x 500 (SPIRE 500 μ m)



PACS 70 μ m l=30 coverage



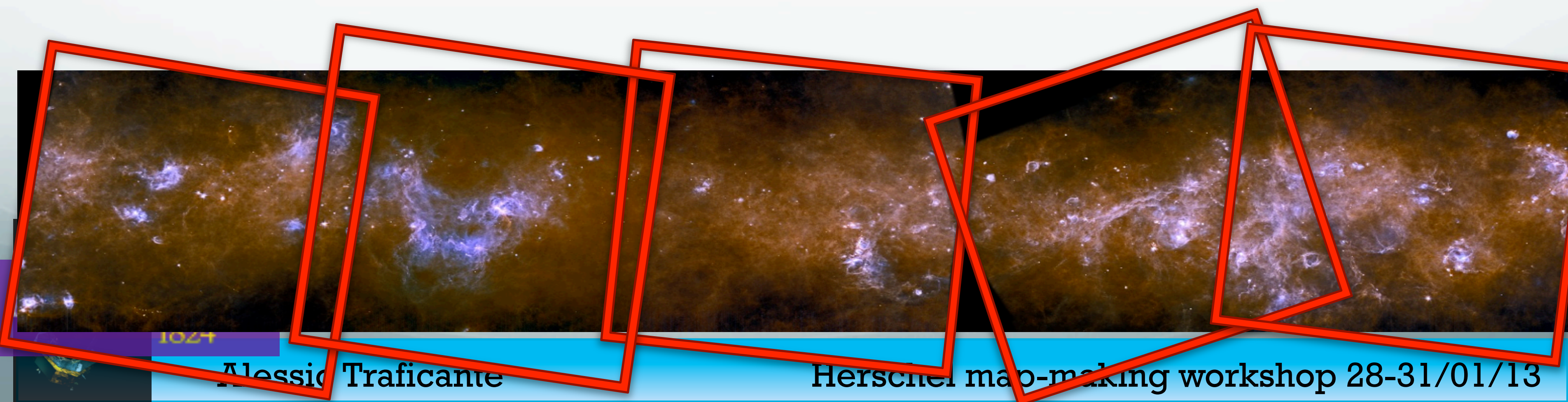
Hi-GAL survey

Hi-GAL <u>Herschel Infrared Galactic Plane Survey</u> (Molinari et al. 2010)

More than 160 2° x 2° tiles

PACS band (μm)	On-board freq. (Hz)	Data freq. (Hz)	Data per tile
70	40	5	~8 Gb
160	20	5	~4 Gb

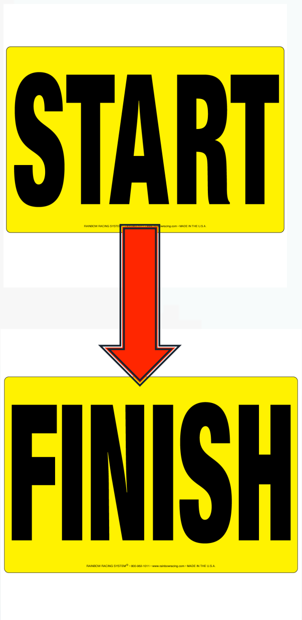
SPIRE band (μm)	Data freq. (Hz)	Data per tile
250	10	~0.8 Gb
350	10	<0.5 Gb
500	10	< 0.5 Gb



MapMaking

ROMAGAL tailored and developed for Hi-GAL "l"

63 tiles $|b| < 1°$ $-71° < l < 66°$



RAW data

Dimensions	**~ Gb**
Contents	Signal+noise (systematics, glitches, statistical fluctuations, …)

Map

Dimensions	**~ Mb**
Contents	Signal + residuals, ideally noiseless

How to reduce data without loosing information? **Map Making**

MapMaking

For each Herschel bolometer, we can model the time ordered data (TOD) **d:**

$$\mathbf{d} = P\mathbf{m} + \mathbf{n}$$

Very simple **IF** beam is circularly symmetric

Pointing Matrix

Map to estimate

Noise

Modeling the noise $\mathbf{n}$ is the key to estimate $\mathbf{m}$

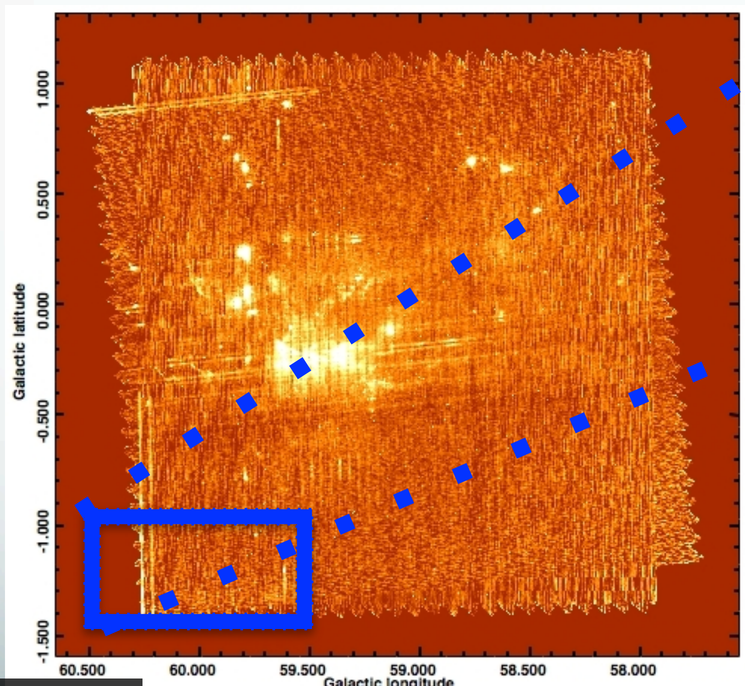
$\mathbf{n}$ depends from
- Technical specification of the instruments
- Specifics of each survey (acquisition strategy, …)

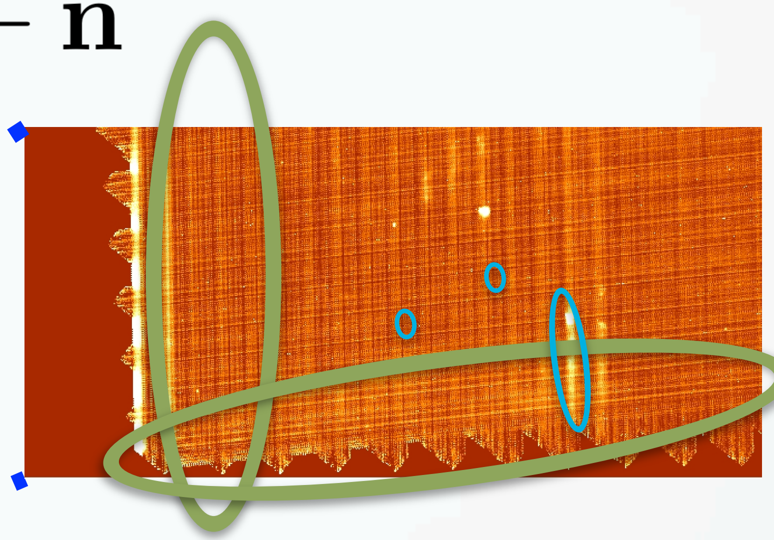
Understanding the noise: naïve map

$$\mathbf{d} = P\mathbf{m} + \mathbf{n}$$

$\mathbf{n}$ → White and stationary noise

Naïve solution $\tilde{\mathbf{m}} = (\mathrm{P}^T\mathrm{P})^{-1}\mathrm{P}^T\mathbf{d}$





1. Systematics (drifts, …)
2. Glitches
3. White noise
4. Correlated noise

l=59° Hi-GAL
PACS 70 μm

ROMAGAL pre-processing

ROMAGAL algorithm

ROMAGAL pre-processing

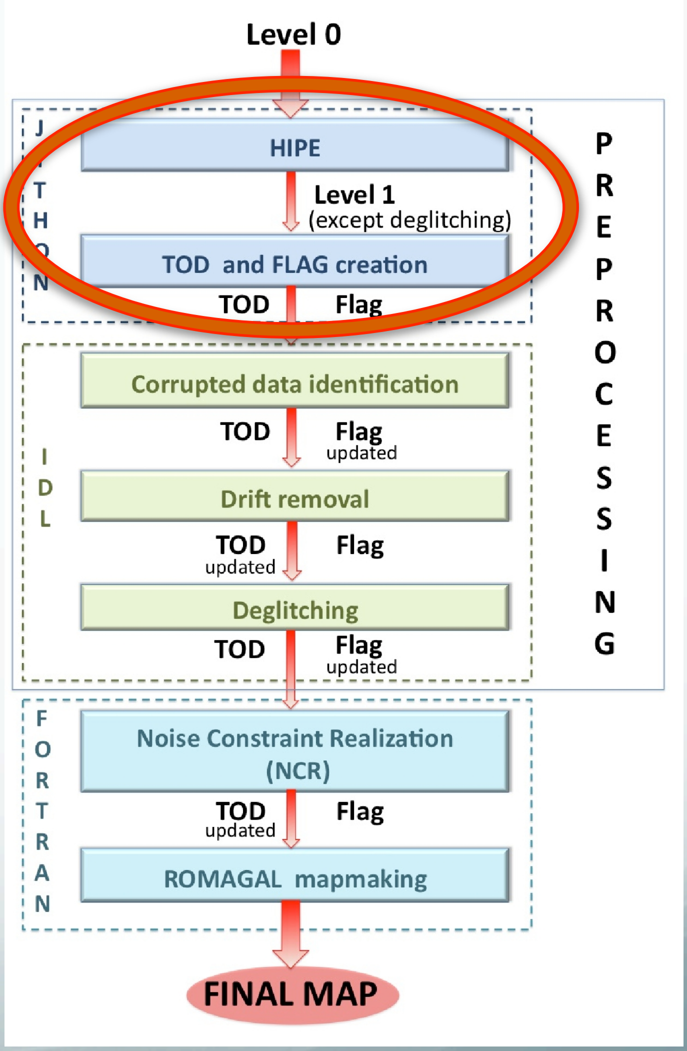
From Level 0 to Level 1: HIPE

- Conversion from ADU to physical unit

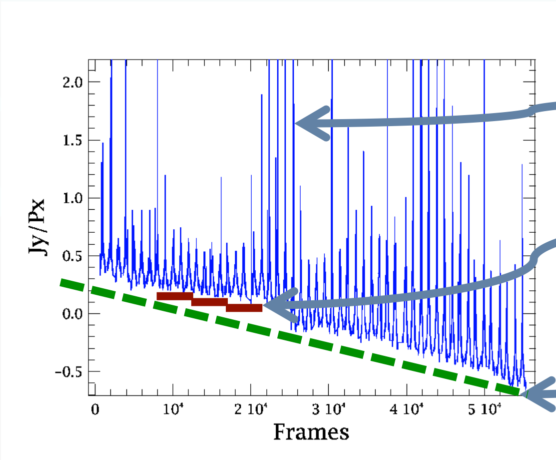
PACS Jy/px SPIRE Jy/beam

- Pixel-to-pixel offset
- …

- **TOD creation for each PACS/SPIRE bolometer**



ROMAGAL pre-processing



Signal

Scan-leg

Drift

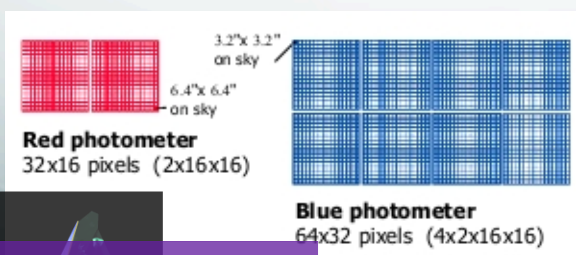
PACS 70 μm TOD

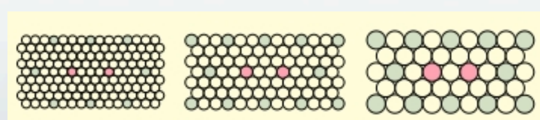
DRIFTS

- telescope drift
- electronic drift

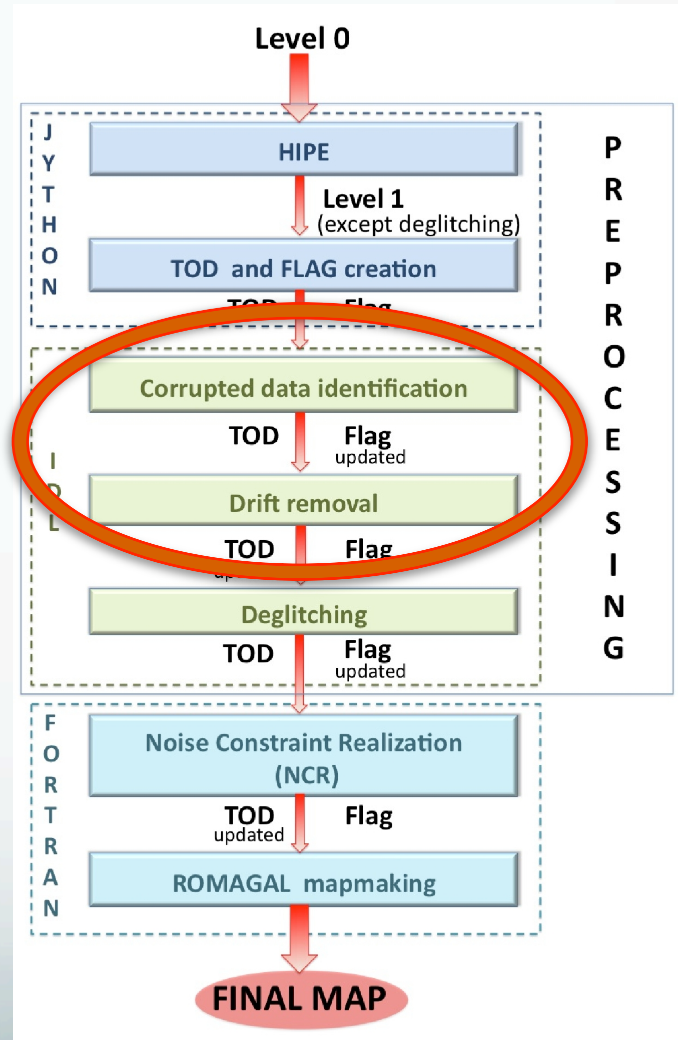
Common to the entire array

Different for each sub-array

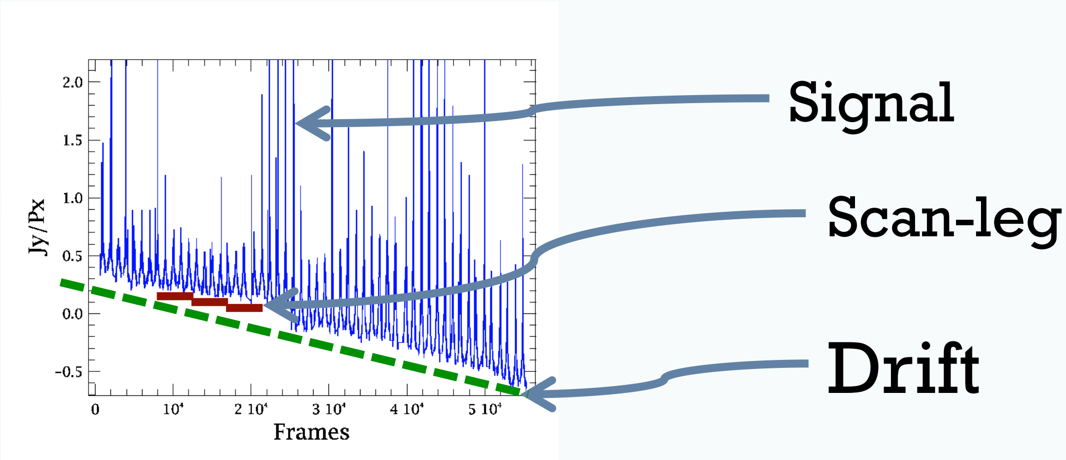




250 μm 350 μm 500 μm



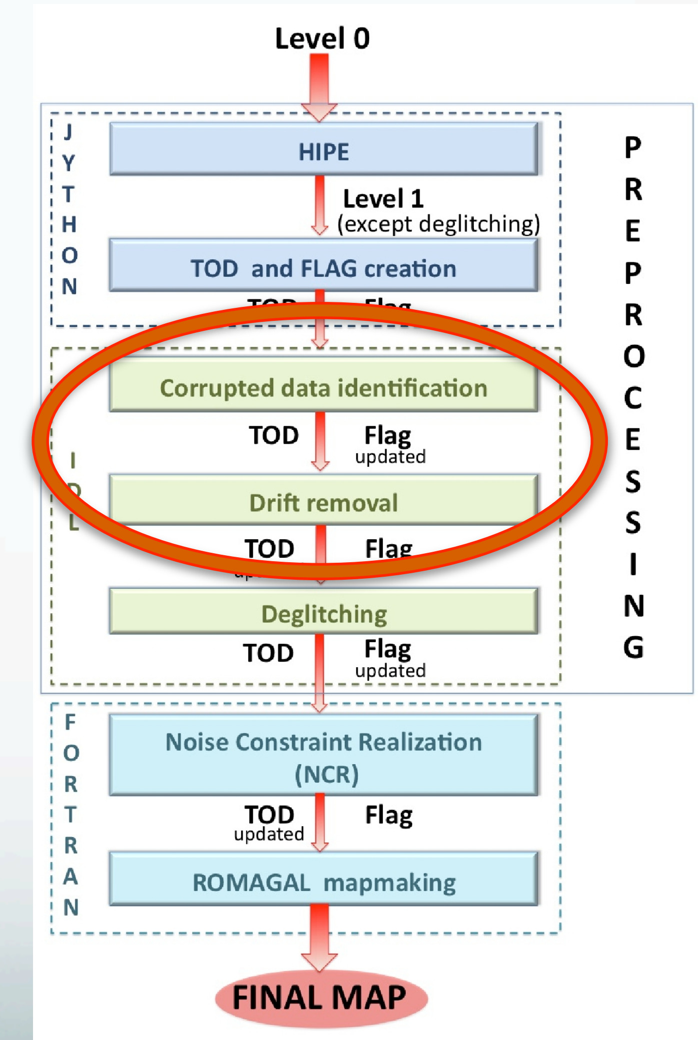
ROMAGAL pre-processing



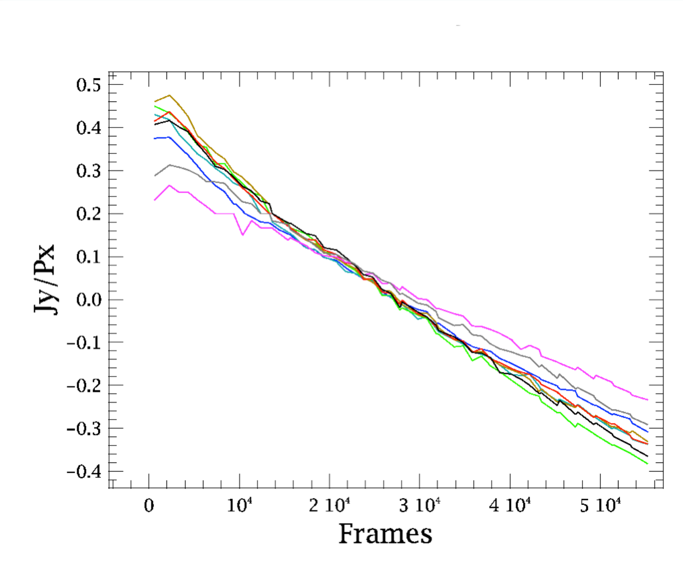
PACS 70 μm TOD

For each sub-array

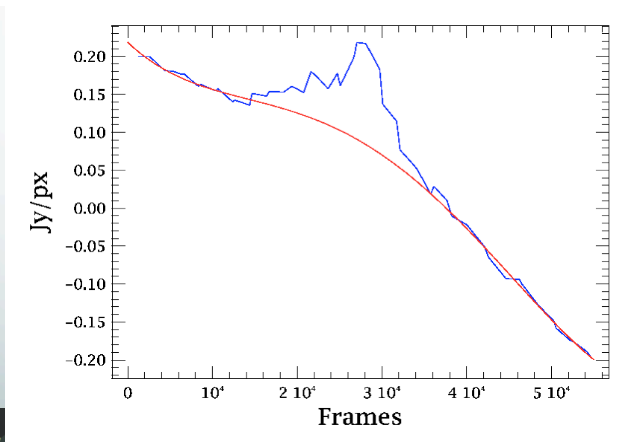
- TOD median value of each scan-leg
- Minimum median as representative of the drift
- Polynomial fit



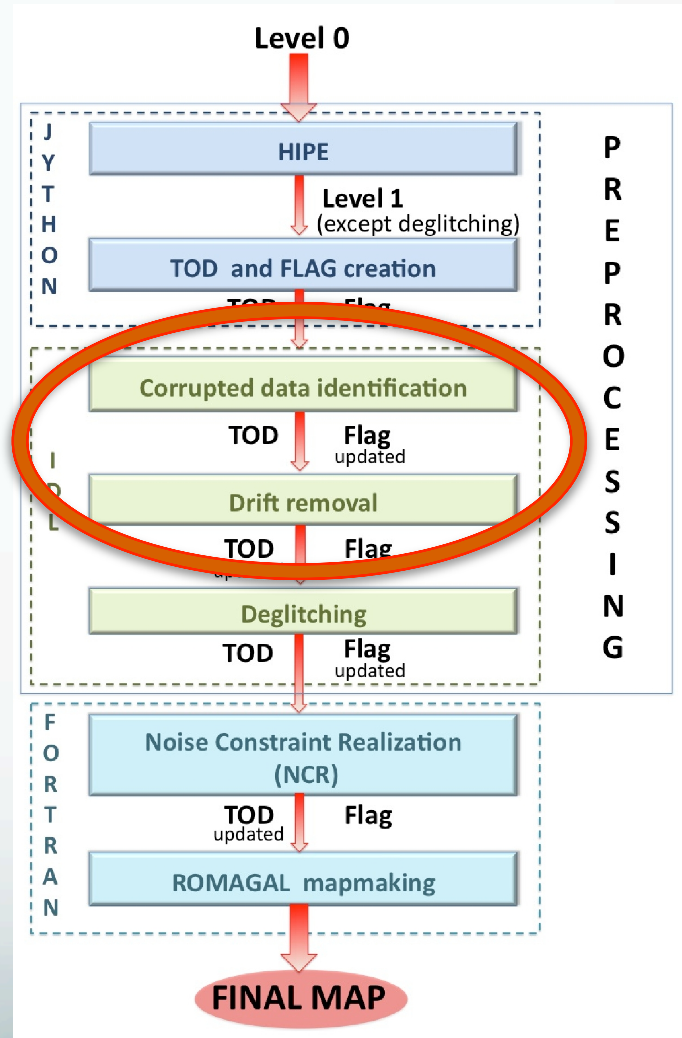
ROMAGAL pre-processing



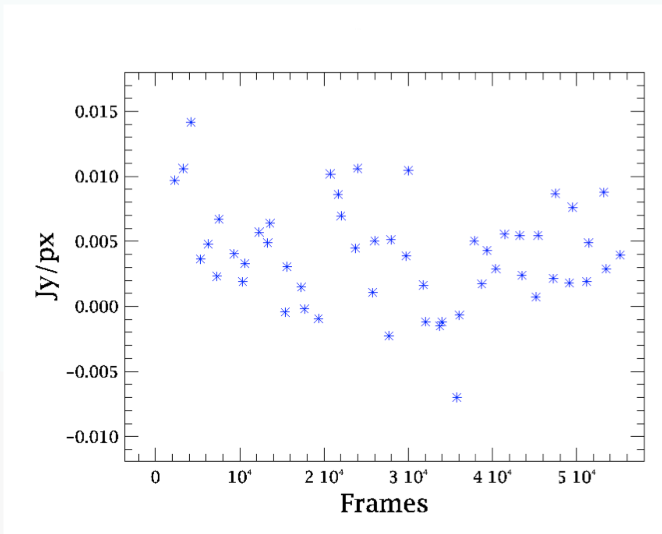
PACS 70 μm minima of median for each sub-array



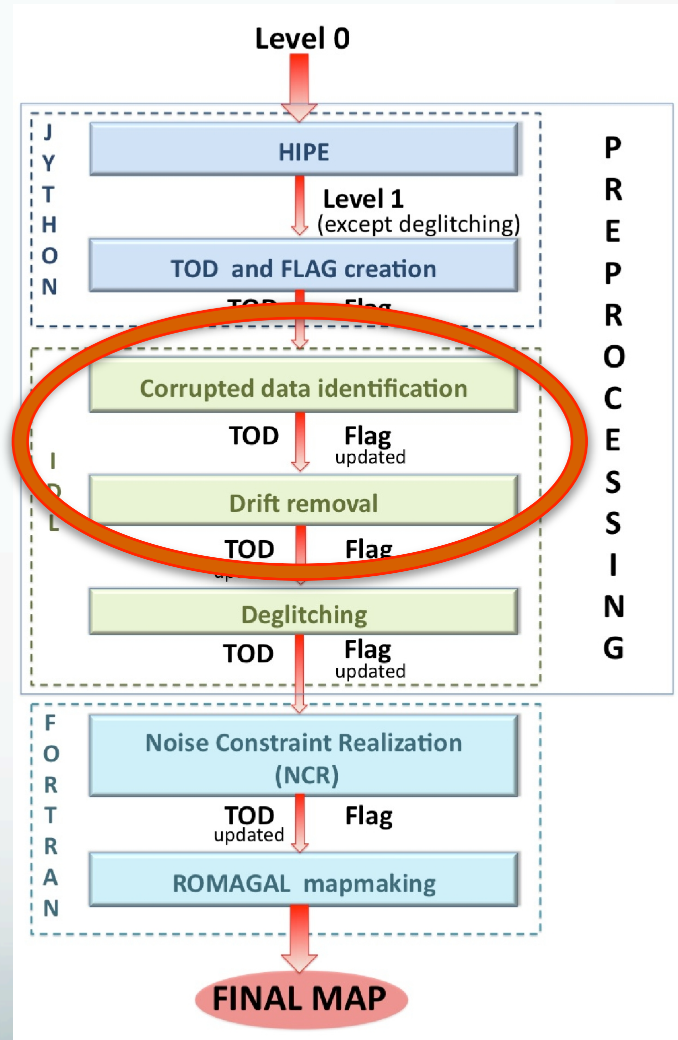
Scan parallel to the Galactic Plane



ROMAGAL pre-processing



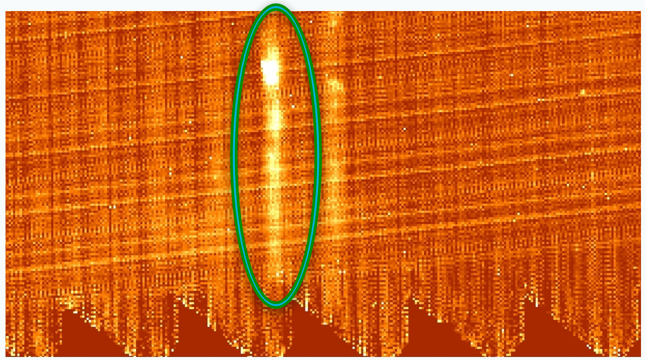
Minimum of the median residual after polynomial fit subtraction



ROMAGAL pre-processing

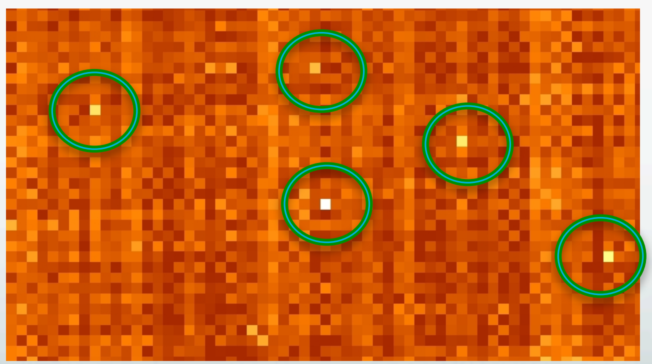
Glitches

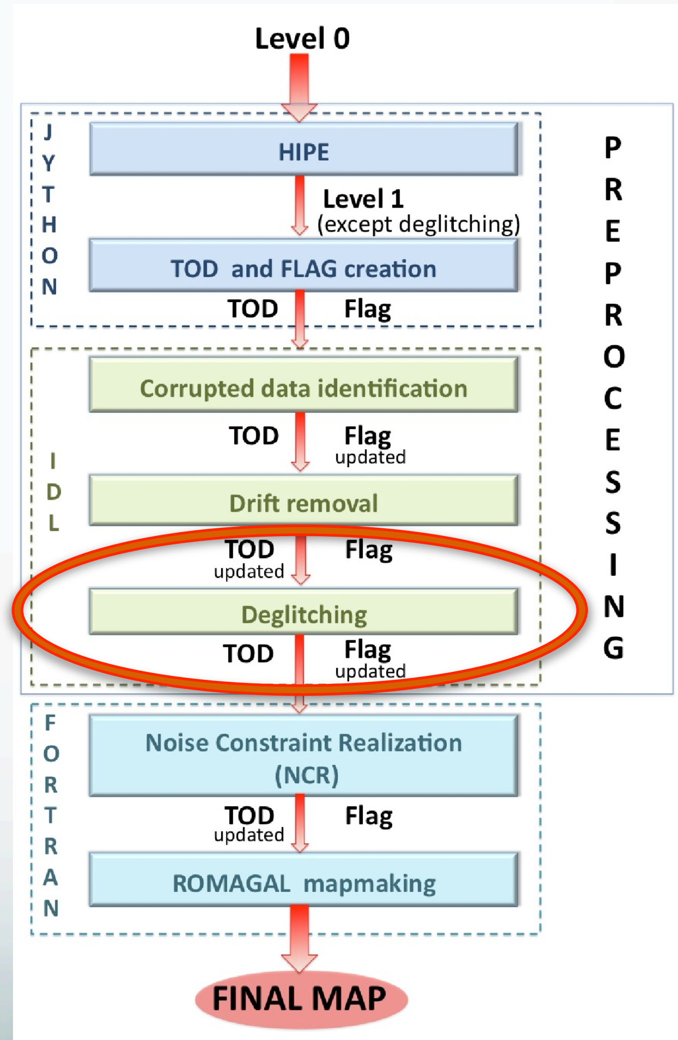
Glitches which alter the detector responsivity



Spikes

Affect only a specific map pixel

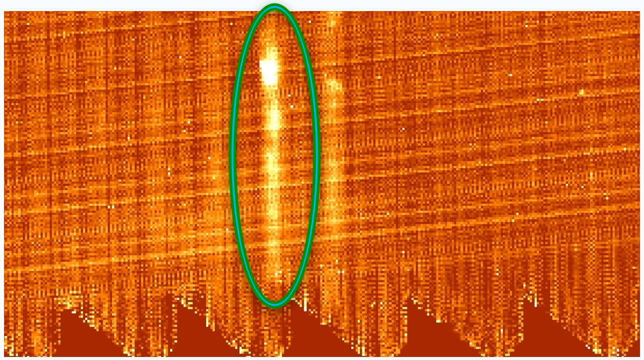


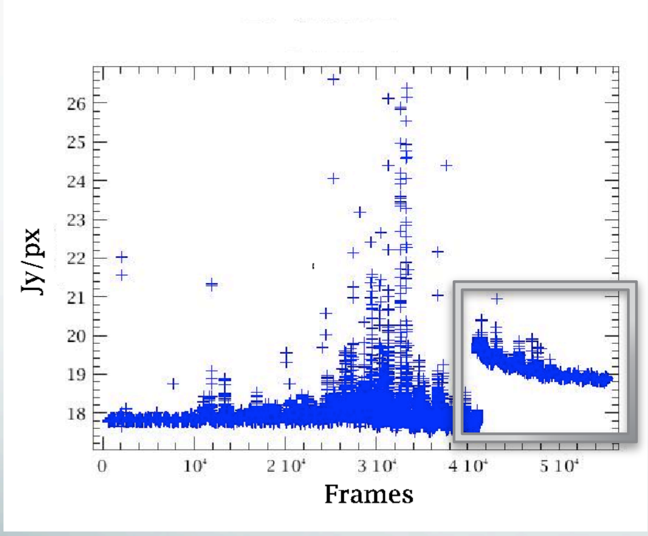


ROMAGAL pre-processing

Glitches

Glitches which alter the detector responsivity

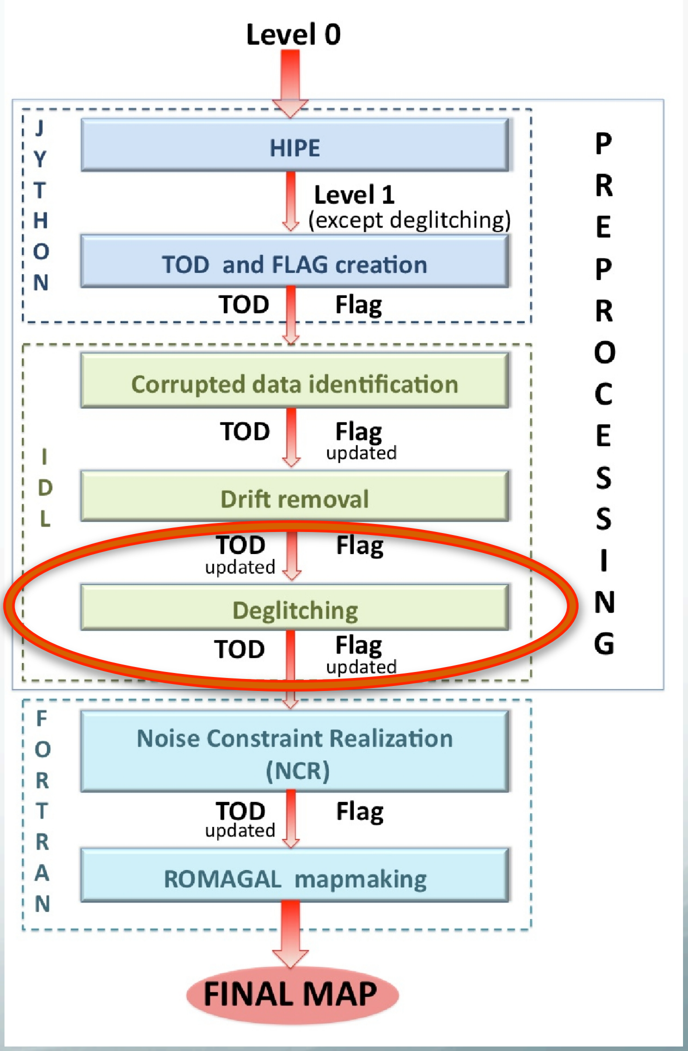




TOD affected by powerful glitch event:

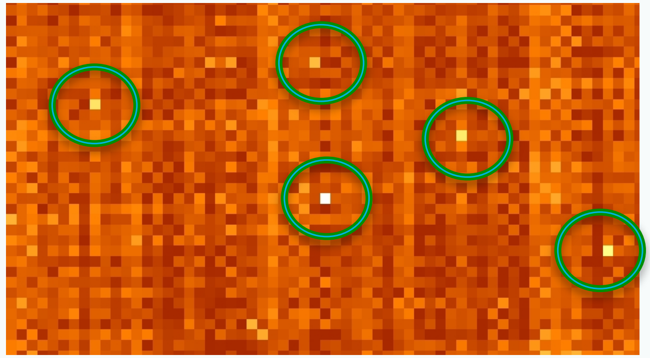
- TOD first derivative to identify the "jump"
- exponential behavior

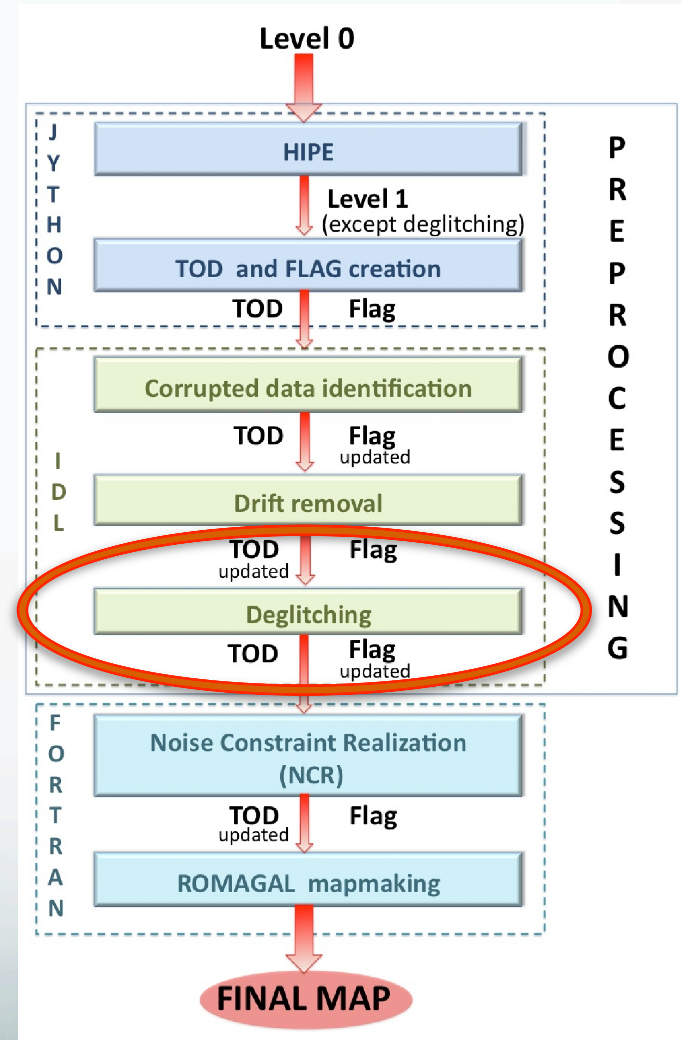
Flagged points



ROMAGAL pre-processing

Spikes





PACS

IDL algorithm

SPIRE

Standard MMT + IDL algorithm

IDL algorithm: sigma clipping on spatial redundancy

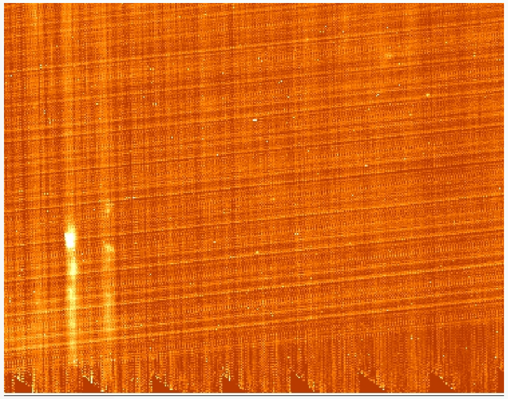
n sigma = $-0.569 + \sqrt{-0.072 + 4.99\log(N)}$

Derived by S. Pezzuto

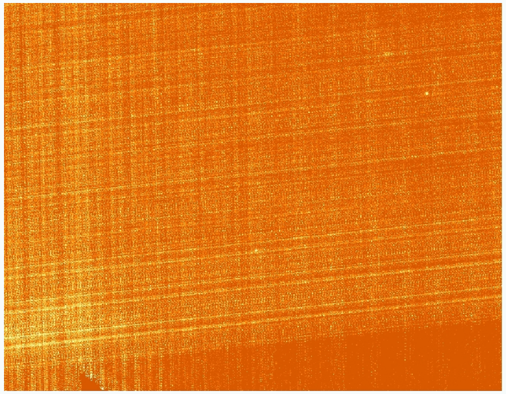
N hits per pixel

~ 5-10% of data **flagged** as glitches

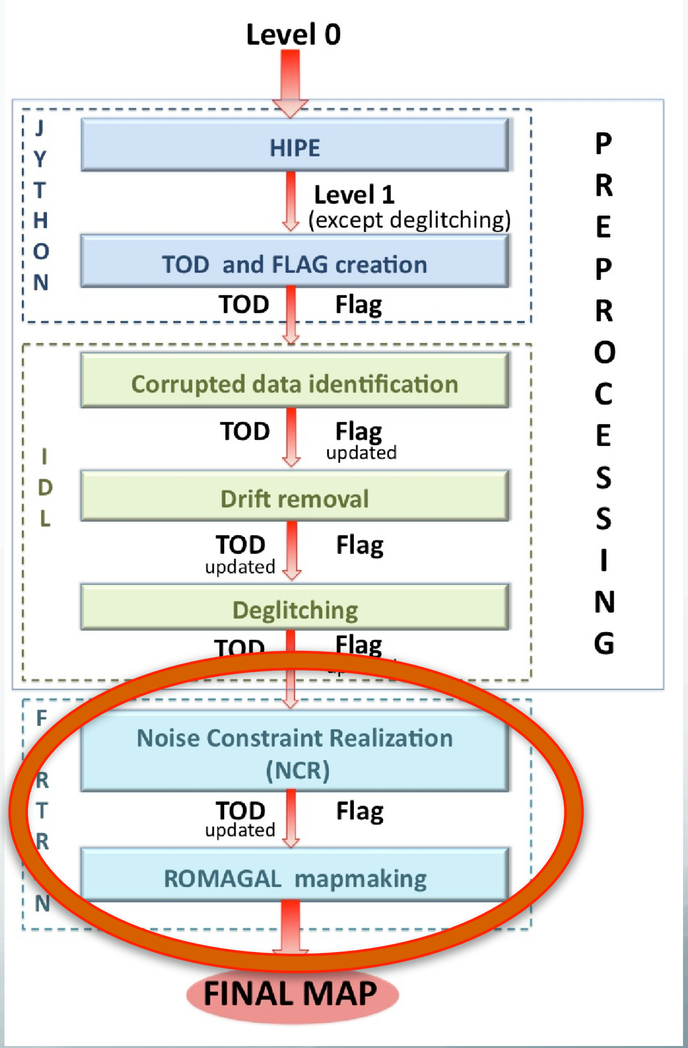
ROMAGAL NCR



PACS 70 μ m naïve
from raw-data



PACS 70 μ m naïve
after pre-processing



Survived: white noise
1/f noise

ROMAGAL
algorithm

But first: how to manage flagged data?

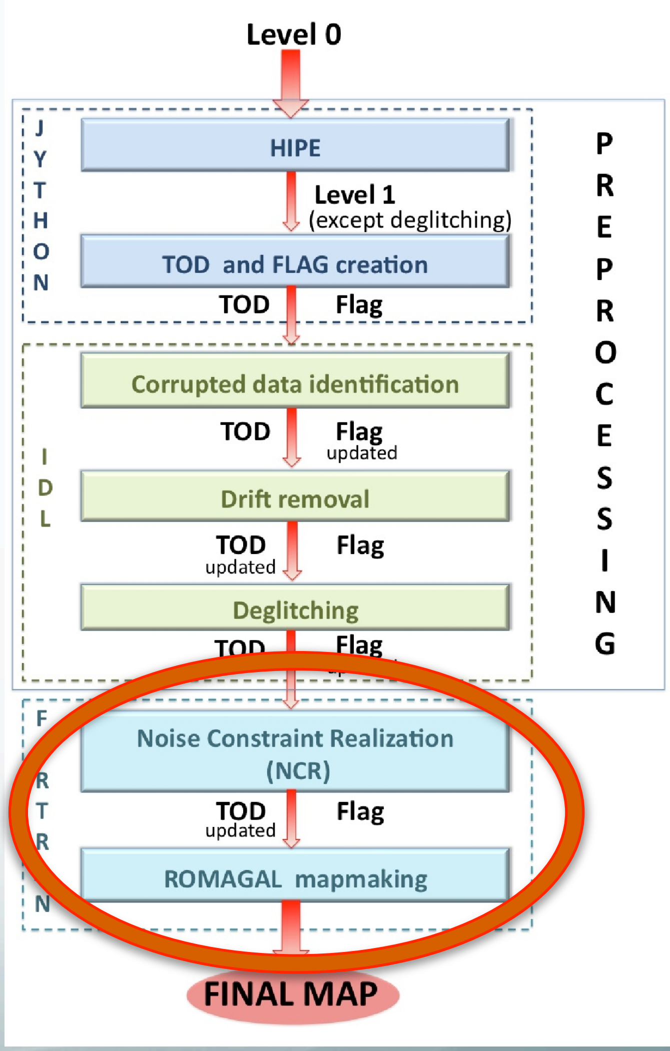
ROMAGAL pipeline

Main ROMAGAL assumption

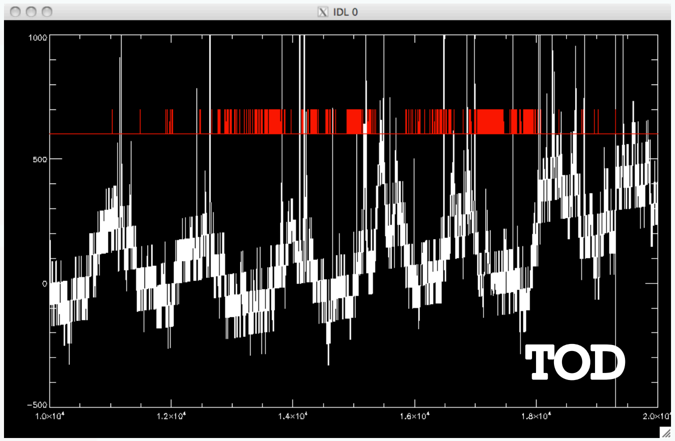
1. The noise **n** is Gaussian distributed and with a null average
2. The noise **n** is piecewise stationary
3. The pure signal does not change with time over the sky

From 1-2 — Noise has to be constrained

From 3 — TODs require continuity



ROMAGAL pipeline



Flagged points

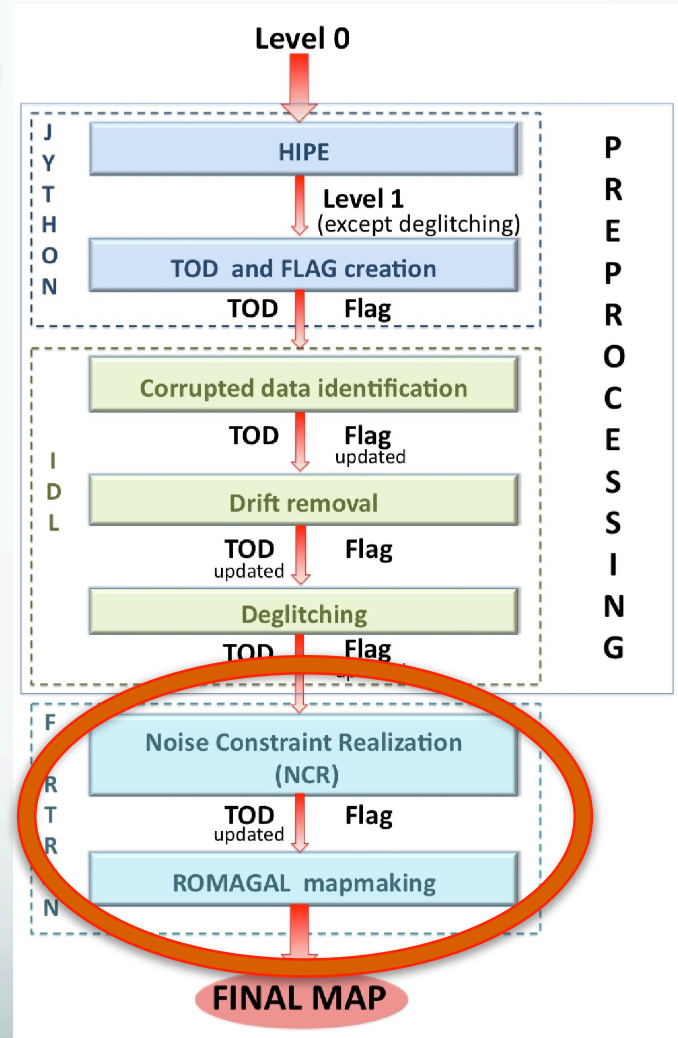
How to substitute the flagged data ?

Simplest solution

Null value

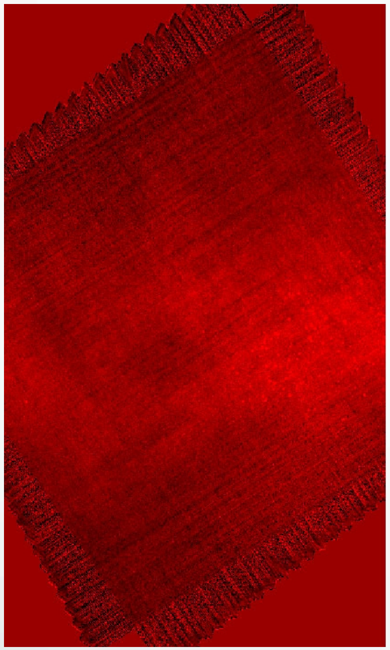
Optimal solution

Noise Constraint Realization

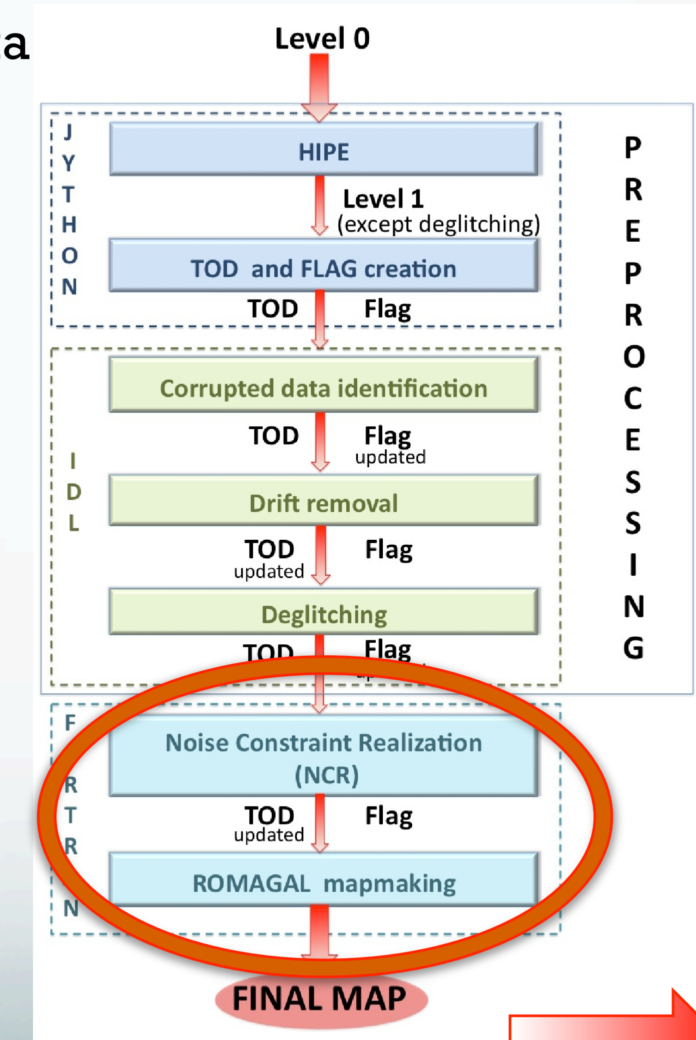


ROMAGAL NCR

Simulations of 1/f noise with 10% of flagged data

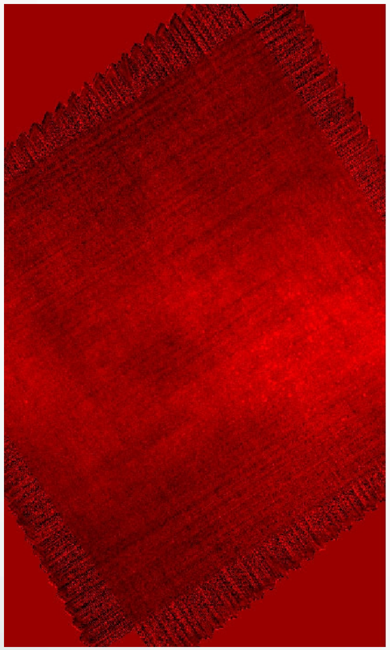


Null value
(Breaks hyp. 1-2-3)

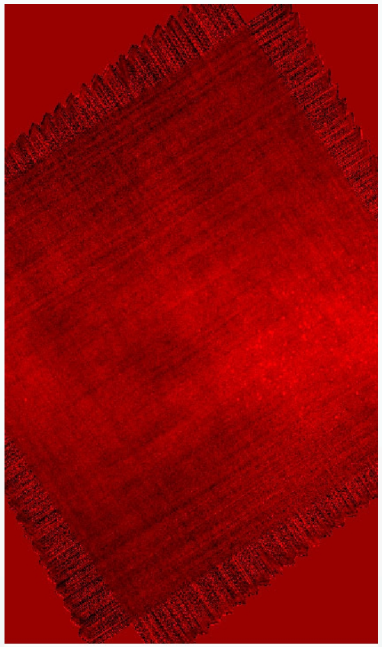


ROMAGAL NCR

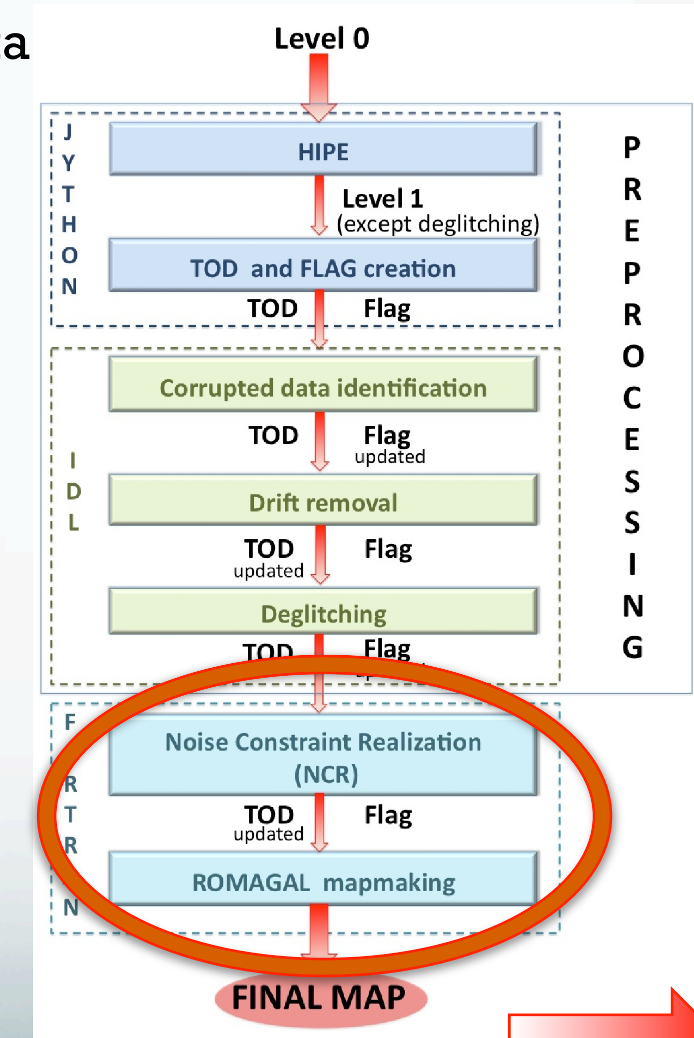
Simulations of 1/f noise with 10% of flagged data



Null value
(Breaks hyp. 1-2-3)

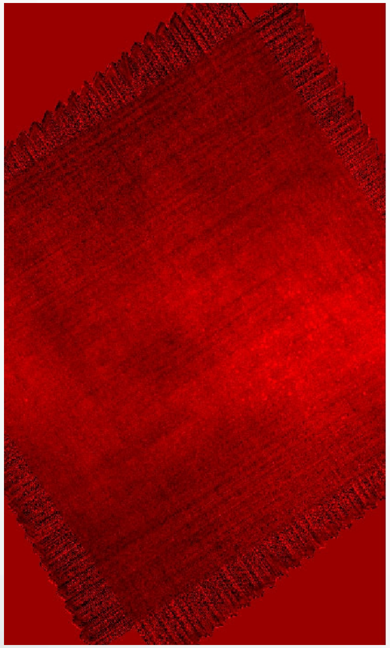


Statistical noise properties preserved (NR)
(Breaks hyp. 3)

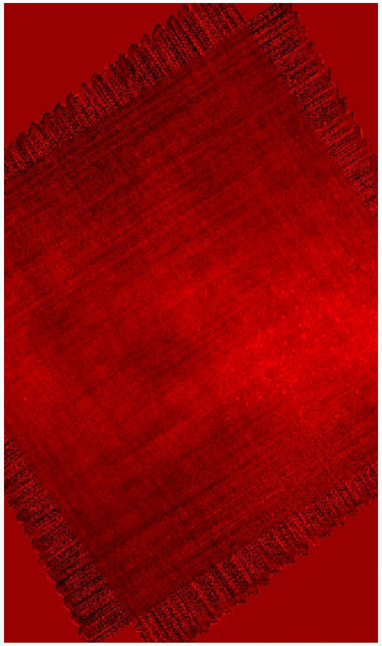


ROMAGAL NCR

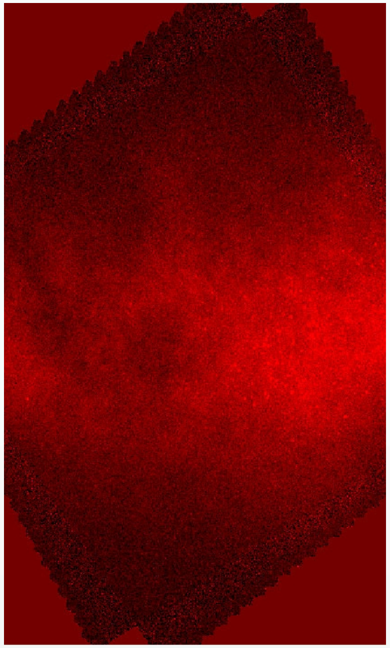
Simulations of 1/f noise with 10% of flagged data



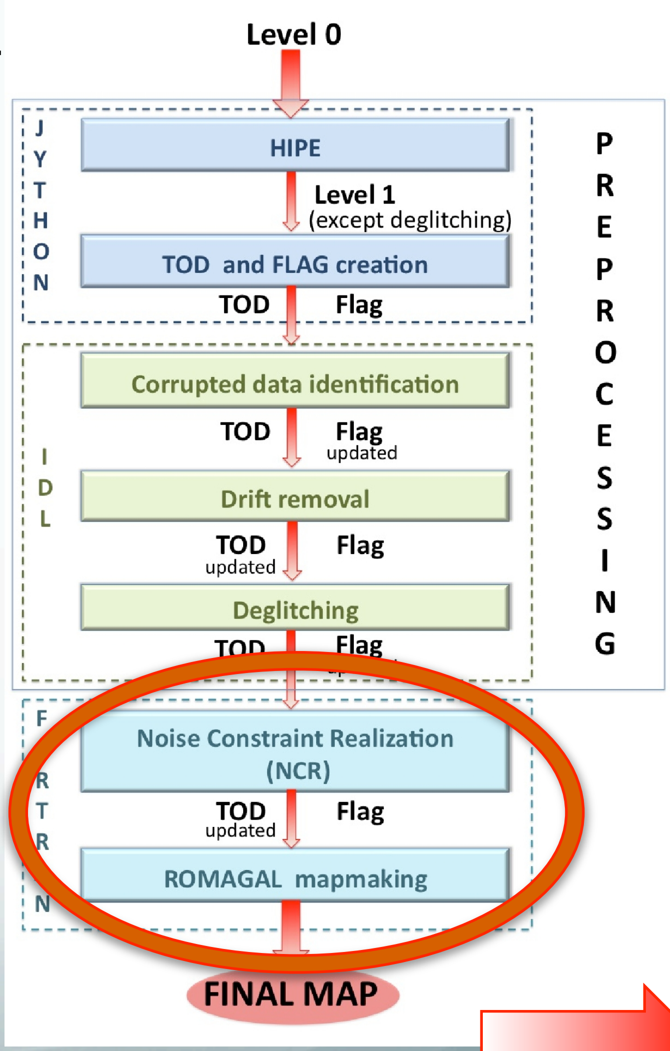
Null value
(Breaks hyp. 1-2-3)



Statistical noise properties preserved (NR)
(Breaks hyp. 3)



Boundary conditions preserved
(NCR)



ROMAGAL MapMaking

- The noise $\mathbf{n}$ is piecewise stationary
- The noise $\mathbf{n}$ is Gaussian distributed and with a null average

$$\langle \mathbf{n} \rangle = 0$$

- The pure signal does not change with time over the sky

We can estimate the best map, solving the system $\mathbf{d} = P\mathbf{m} + \mathbf{n}$ via **GLS** (Generalized Least Square) solution

$$\boxed{\tilde{\mathbf{m}} = (P^T \mathrm{N}^{-1} P)^{-1} P^T \mathrm{N}^{-1} \mathbf{d}}$$

$\mathrm{N} = \langle \mathbf{n}\mathbf{n}^T \rangle$ Noise Correlation matrix (NON-diagonal)

$\tilde{m}$ is
- unbiased
- optimal minimum variance estimator
- Maximum Likelihood estimator

ROMAGAL implementation

GLS solution $\tilde{\mathbf{m}} = (P^T \mathrm{N}^{-1} P)^{-1} P^T \mathrm{N}^{-1} \mathbf{d}$

This matrix is $n \sim 10^7$ Inversion scales as n^3

$n = 10^{21}$ operations requires $\sim$ 100 days…

…on TITAN!!!

N° 1 supercomputer to date
$\sim$ 500.000 processors
$\sim$50 petaflops (10^{15} oper./s) peak performance

We should find some smart workaround….

ROMAGAL implementation

GLS solution $$\tilde{\mathbf{m}} = (P^T \mathrm{N}^{-1} P)^{-1} P^T \mathrm{N}^{-1} \mathbf{d}$$

Real implementation $$(\mathrm{P}^T \mathrm{N}^{-1} \mathrm{P}) \tilde{\mathbf{m}} = \mathrm{P}^T \mathrm{N}^{-1} \mathbf{d}$$

Why? → N is stationary

Noise properties does not change over the observation time

N is ~ **circulant** matrix $$\mathrm{N}_{i+1,j+1} = \mathrm{N}_{i,j}$$

Circulant matrix is **DIAGONAL** in Fourier space

The inversion of a diagonal n×n matrix scales as n !

Natoli et al. 2001
De Gasperis et al. 2005

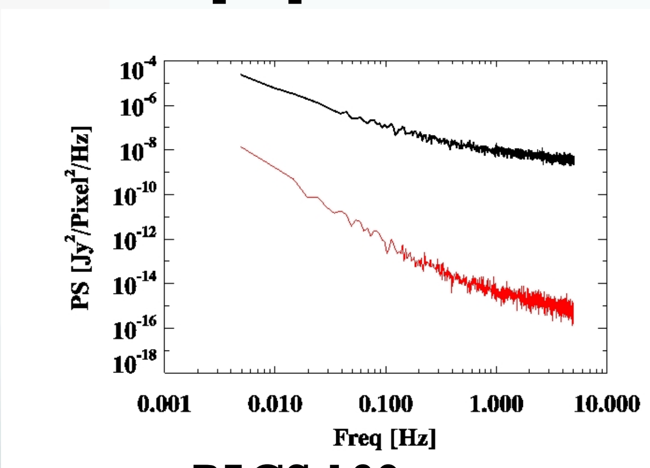
ROMAGAL noise filters

DIRECT space ⟷ Fourier space

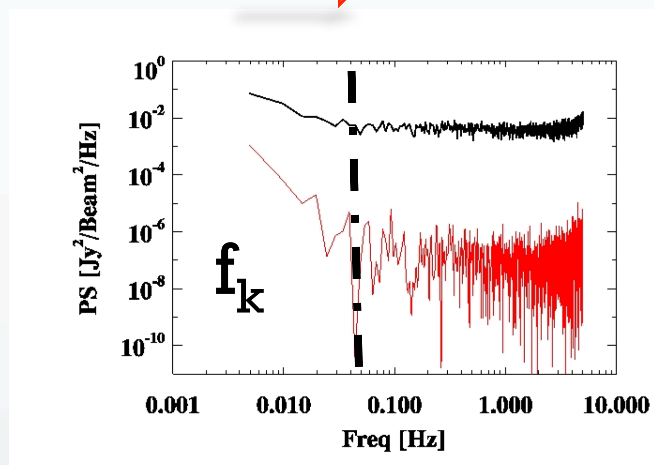
Correlation matrix N ⟷ Noise power spectrum

Inverse of correlation matrix N^{-1} ⟷ Noise filter

Noise filters can be estimated from REAL data. They depend from the intrinsic noise properties of each specific detector → 1 noise filter per bolometer



PACS 160 μm



SPIRE 250 μm

Blank data
(Thanks to G. Pilbratt)

■ Auto spectrum

■ Cross spectrum

f_k = knee frequency   white noise dominate the 1/f noise

ROMAGAL implementation

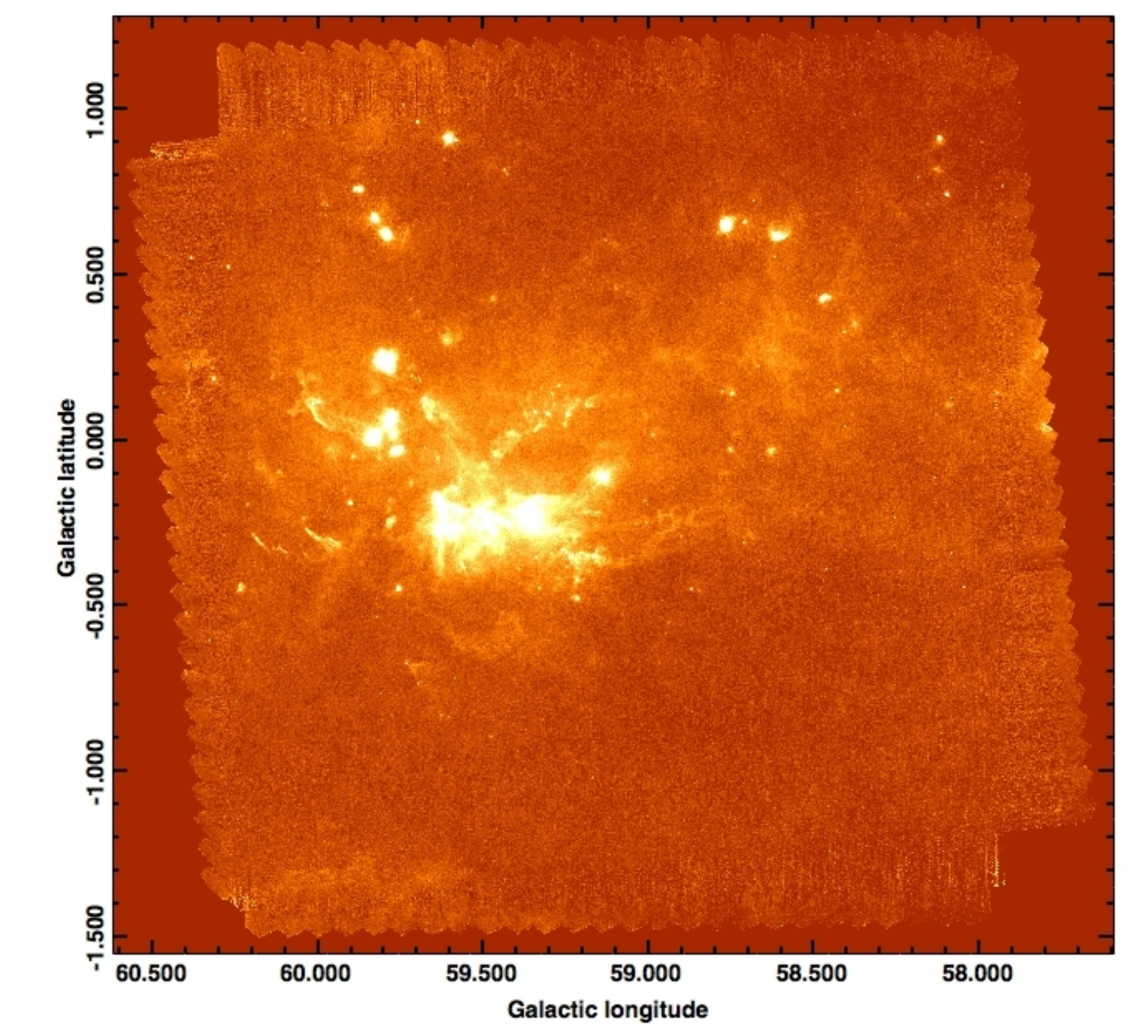
Real implementation $$(\mathrm{P}^T\mathrm{N}^{-1}\mathrm{P})\tilde{\mathbf{m}} = \mathrm{P}^T\mathrm{N}^{-1}\mathbf{d}$$

- Parallel FORTRAN 90/95 code based on MPI libraries
- FFT routines
- Iterative methods (Coniugate Gradient - CG) that converges in few iterations to solve the system

PACS 70 μm map of a 2 ° × 2 ° Hi-GAL field:

~ 20 mins on 8 x 2.5 GHz cores

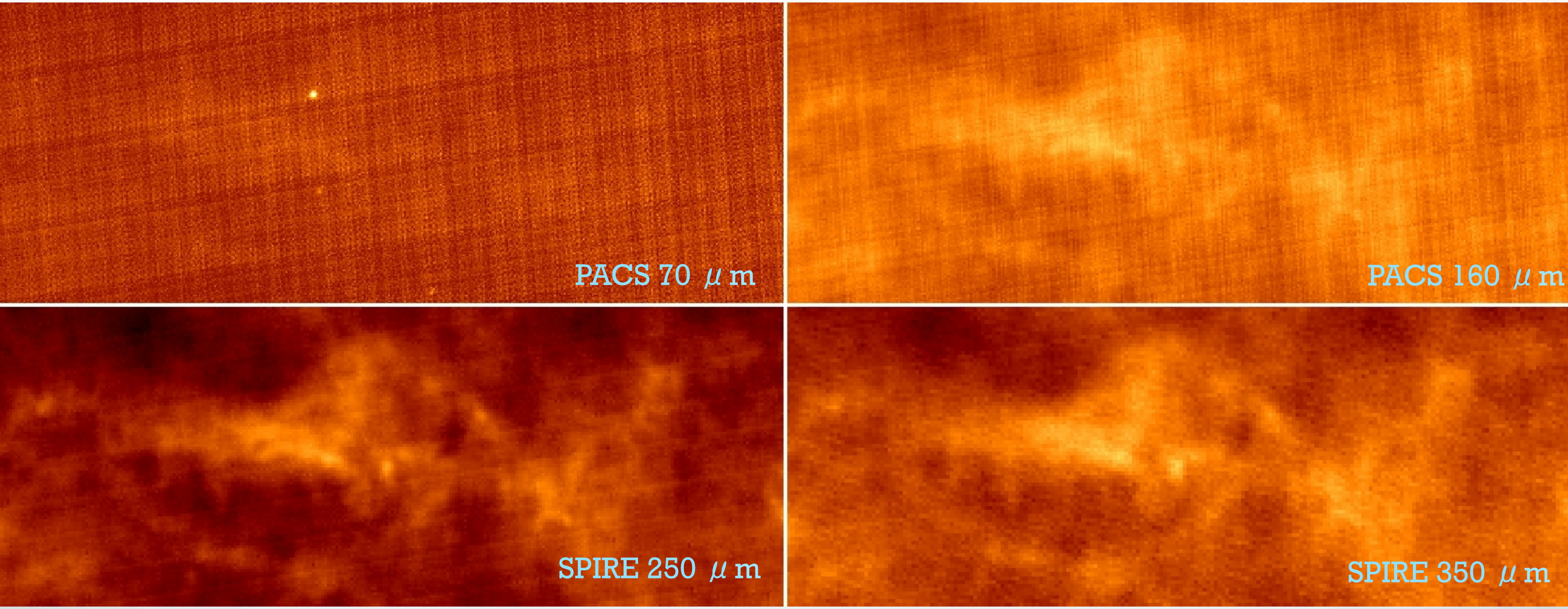
ROMAGAL map



Hi-GAL (l,b)=(59°,0°) PACS 70 μm

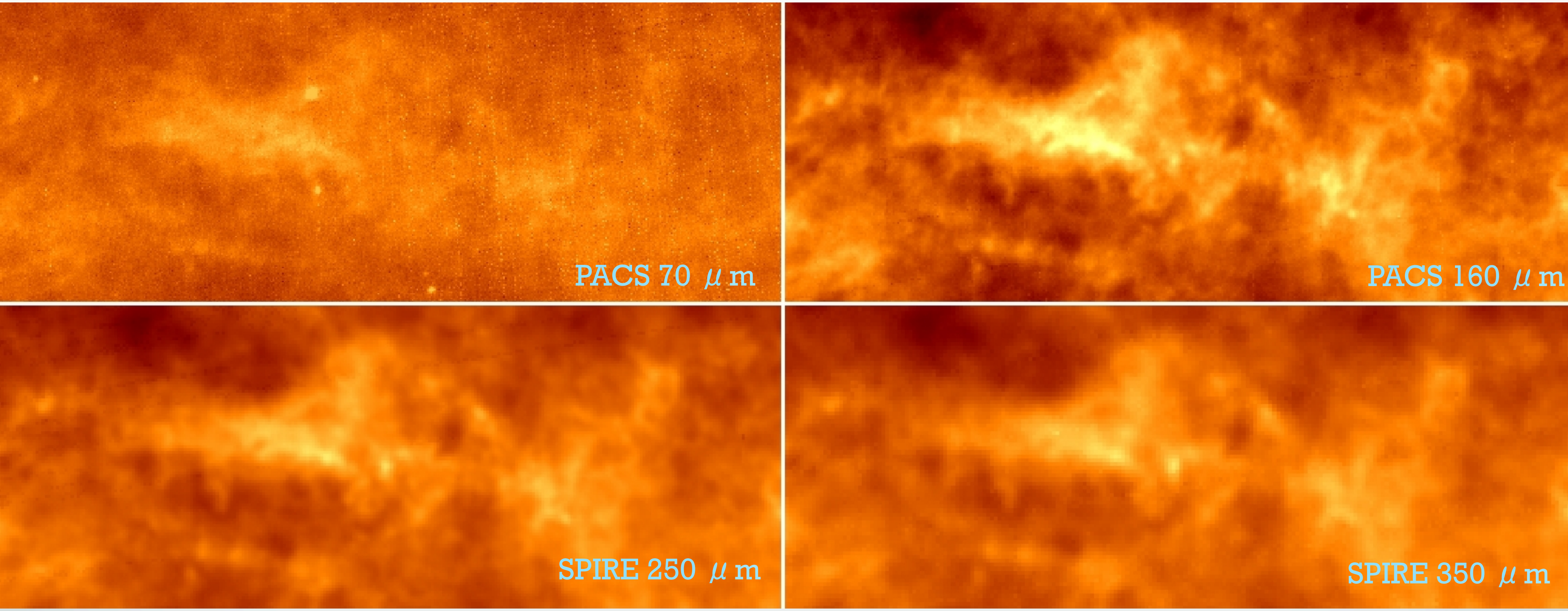
ROMAGAL map

Hi-GAL l=030 naïve



ROMAGAL map

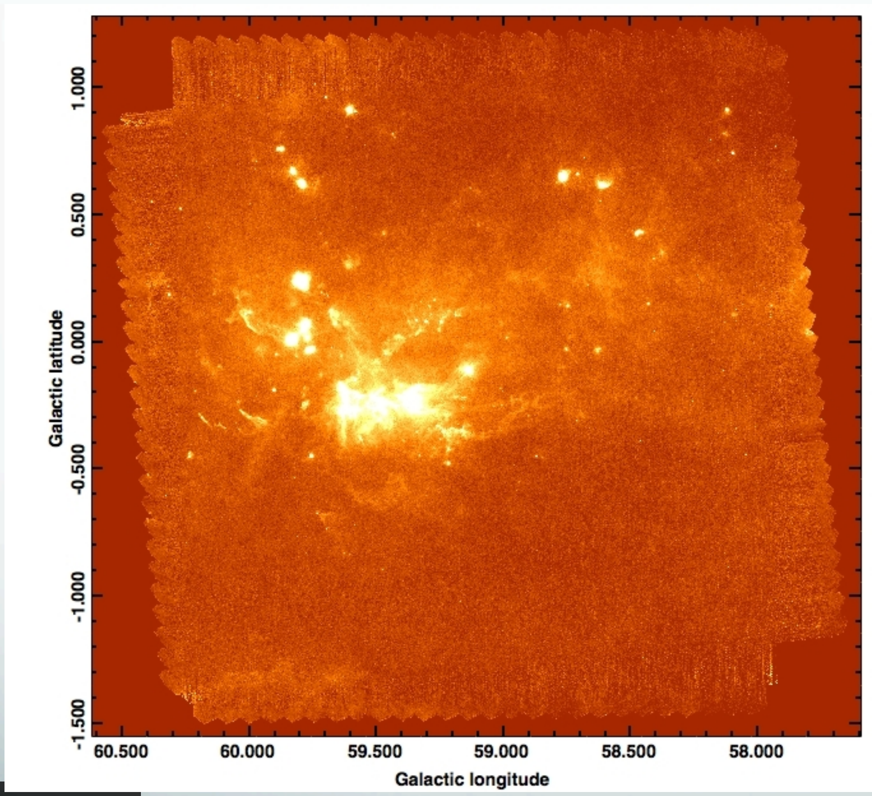
Hi-GAL l=030 GLS



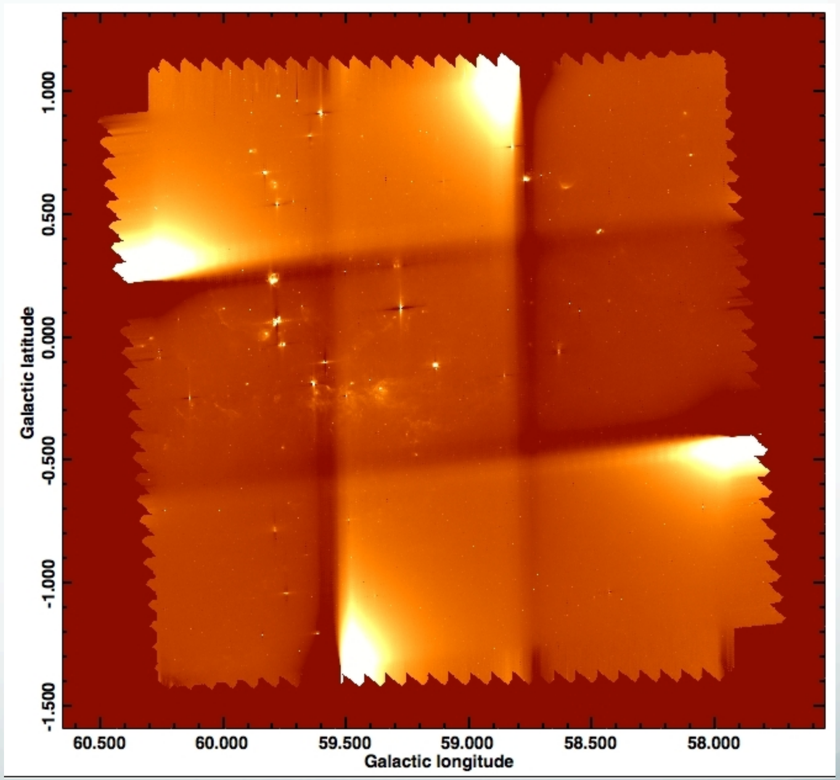
ROMAGAL map

Pre-processing is DEMANDING

WITH pre-processing



WITHOUT pre-processing



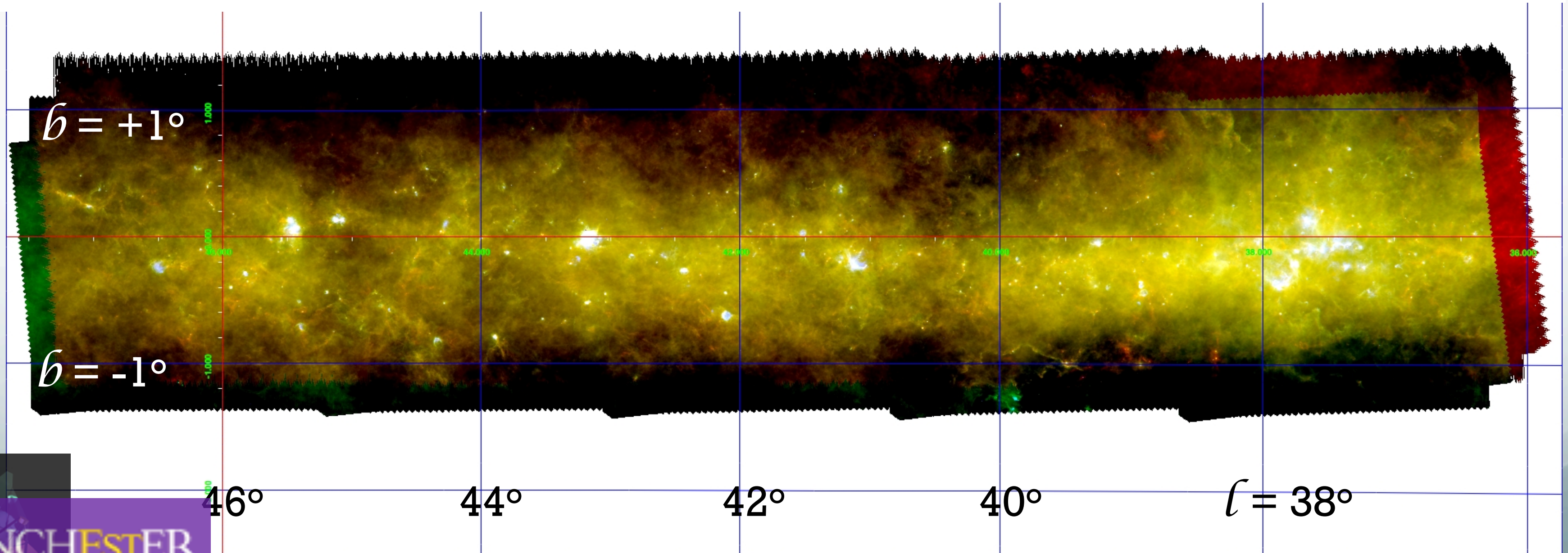
ROMAGAL map

PACS $l = 30°$ field			
Band	rms igls (Jy/pixel)	rms naive(Jy/pixel)	ratio
70μm	0.0085	0.026	~ 3.1
160μm	0.047	0.102	~ 2.2

PACS $l = 59°$ field			
Band	rms igls (Jy/pixel)	rms naive(Jy/pixel)	ratio
70μm	0.004545	0.02208	~ 4,9
160μm	0.01899	0.03586	~ 1,9

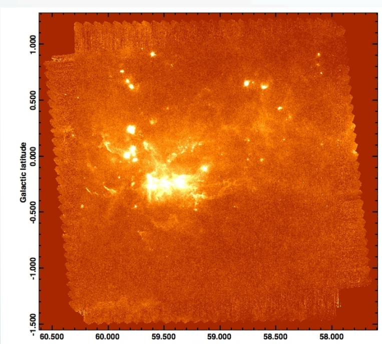
SPIRE $l = 30°$ field			
Band	rms igls (Jy/beam)	rms naive(Jy/beam)	ratio
250μm	0.1749	0.2868	~ 1.6
350μm	0.1569	0.2302	~ 1.5
500μm	0.2659	0.4065	~ 1.5

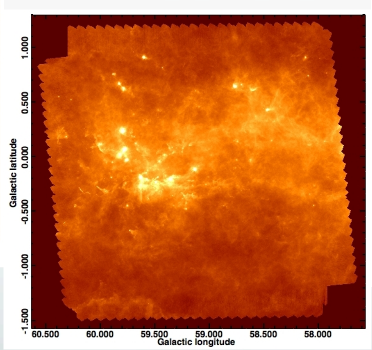
SPIRE $l = 59°$ field			
Band	rms igls (Jy/beam)	rms naive(Jy/beam)	ratio
250μm	0.09857	0.1123	~ 1.1
350μm	0.0734	0.08164	~ 1.1
500μm	0.1073	0.2101	~ 1,9

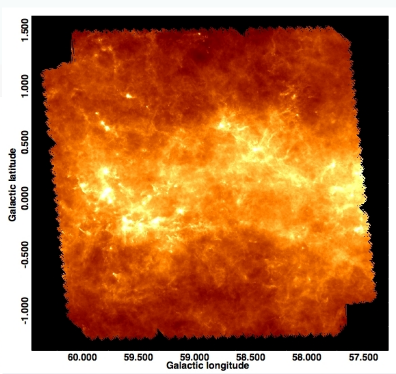


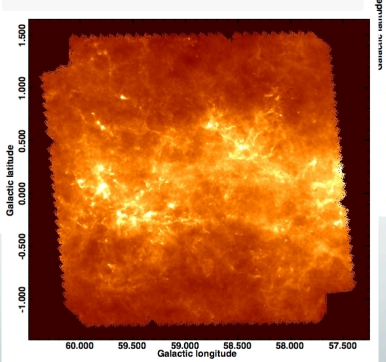
ROMAGAL map

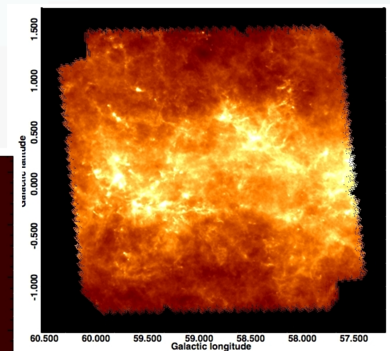
Band	Nominal beam (“)	Pixel size (“)	Tot pixel
PACS 70	5.2	3.2	~ 1700x1700
PACS 160	12.0	4.5	~ 1200 x 1200
SPIRE 250	18.0	6.0	~ 1000 x 1000
SPIRE 350	24.0	8.0	~ 700 x 700
SPIRE 500	34.5	11.5	~ 500 x 500


70 μm


160 μm

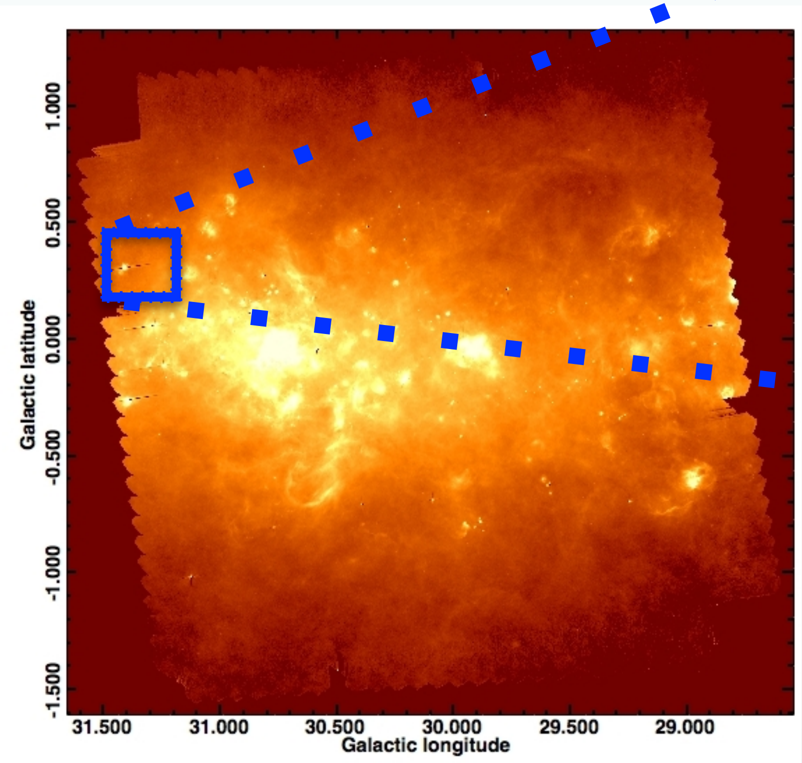

250 μm

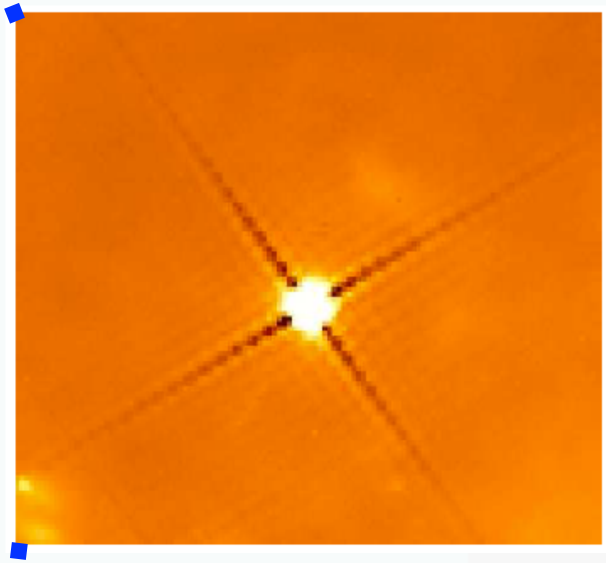

350 μm


500 μm

ROMAGAL map

…but still some problems





Pixellitation noise

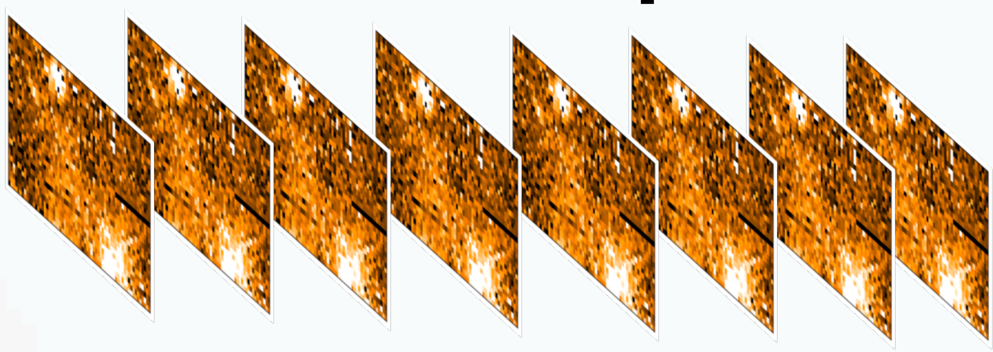
PACS 70 μ m

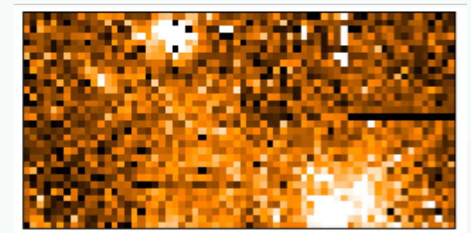
- Time bolometer response
- Relative pointing error
- **Coaddiction**
- …

Co-addiction

PACS 70 μ m on-board sample: 40 Hz

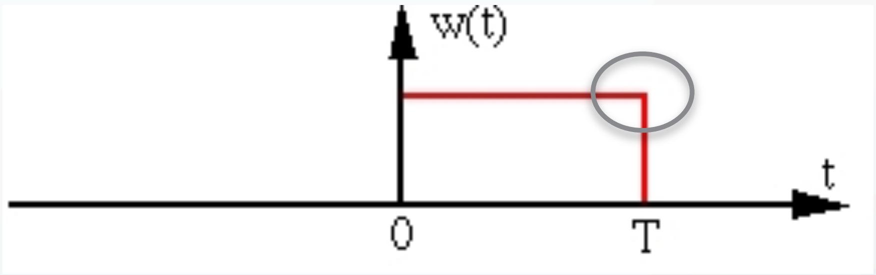
PACS 70 μ m data: 5 Hz



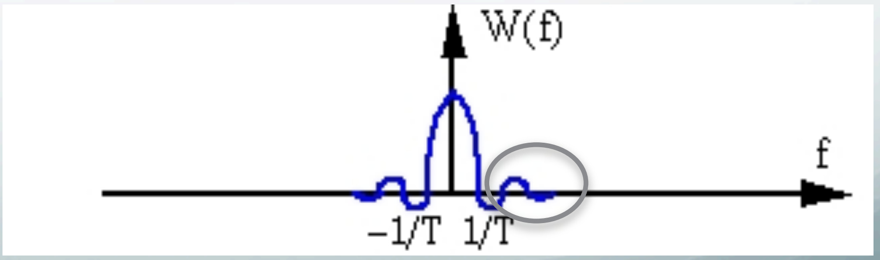


This is a (one of the) problem for ROMAGAL and all Fourier-based algorithms :

Co-addiction in time-space : convolution of TOD with a box-like function

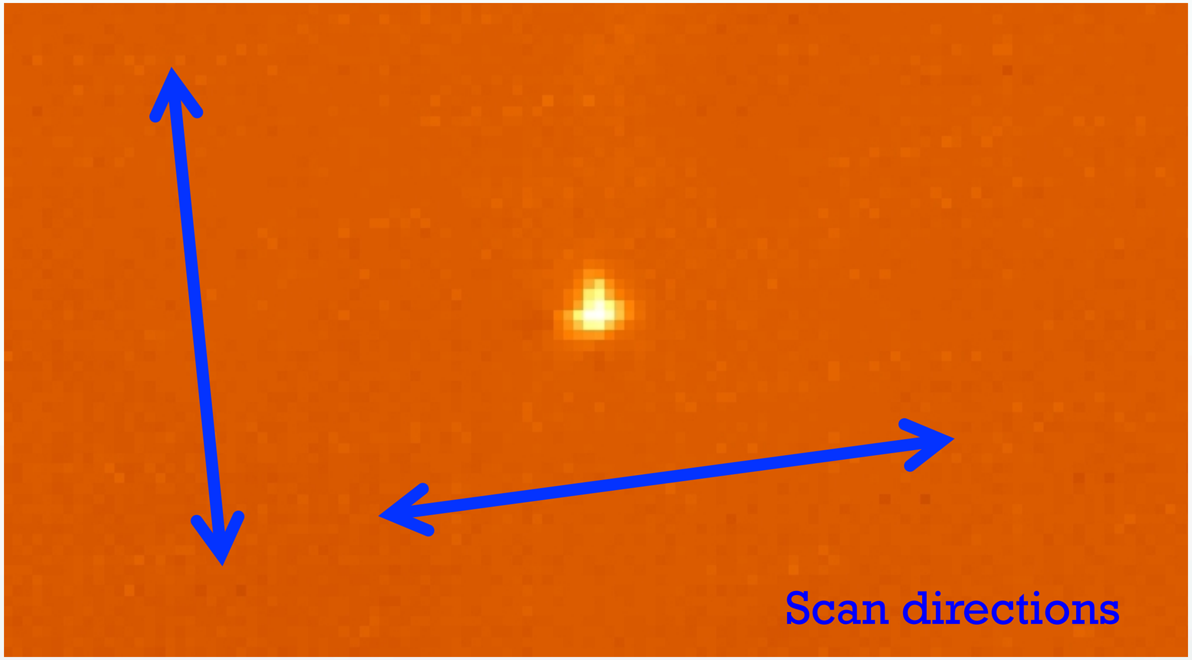


Box function in Fourier space: RINGING!!!



Scan-speed

Beam is highly scan-direction dependent



Beam distortion breaks GLS assumption (pointing matrix P)

ROMAGAL PGLS

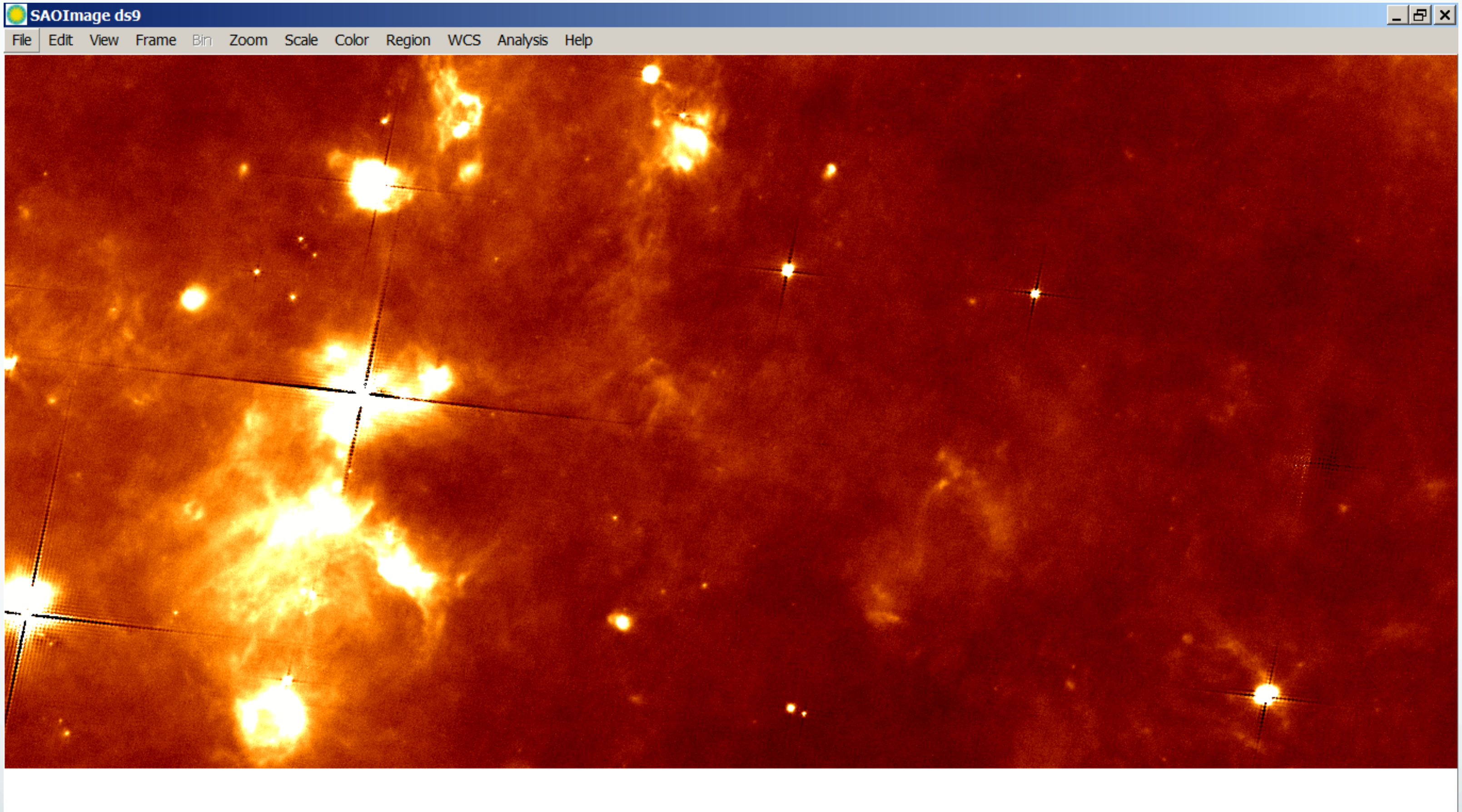
Cannot (?) be cured in a pre-process step

Solution: post-process artifacts analysis and removal PGLS/WGLS (MATLAB)
Piazzo et al. 2012

Post Processed GLS (PGLS) is based on the following artifacts estimation step

1. Unroll the GLS map: $$d_u = P^T \tilde{m}$$
2. Remove signal: $$d_n = d_u - d$$
3. Remove correlated noise: $$d_w = M(d_n)$$
4. Estimate artifacts: $$m_a = P d_w$$

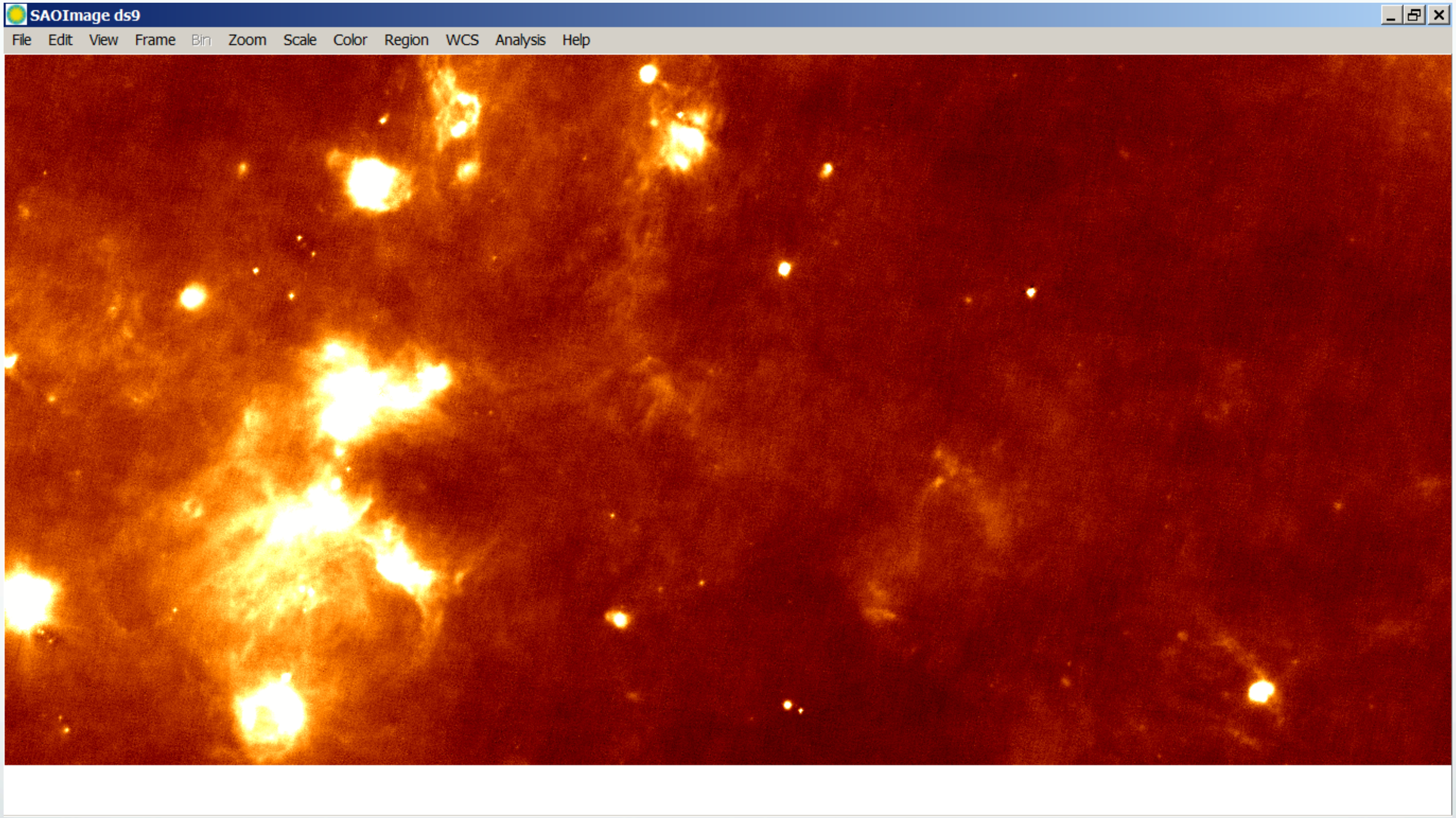
ROMAGAL PGLS



Hi-GAL l323 PACS 70 μ m ROMAGAL

Credits: L. Piazzo

ROMAGAL PGLS



Hi-GAL l323 PACS 70 μm ROMAGAL PGLS

Credits: L. Piazzo

ROMAGAL PGLS - WGLS

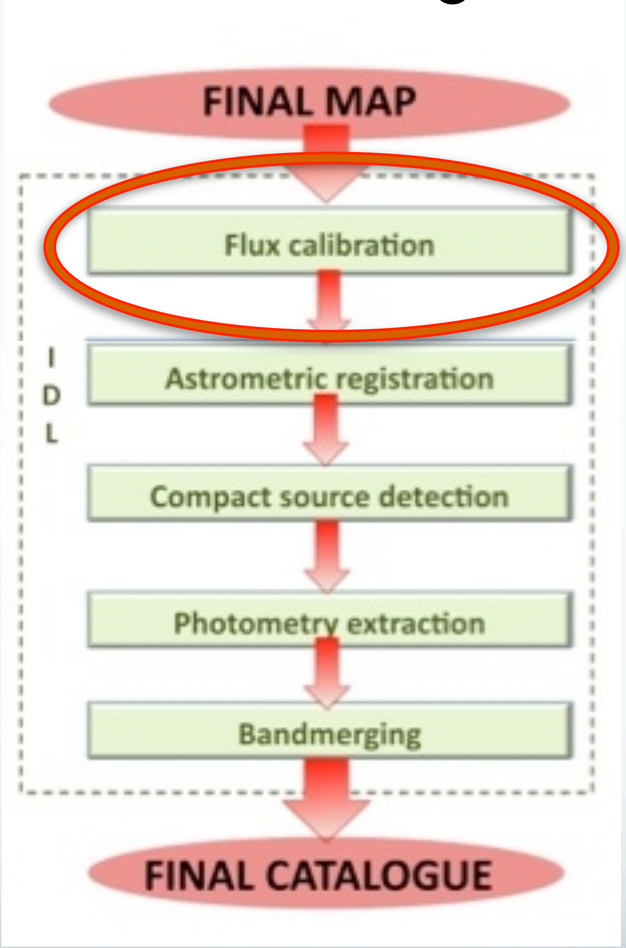
Hi-GAL l323 PACS 70 μm artifacts

WGLS only adds noise where it is negligible (strong sources).
WGLS maps have the same SNR as GLS maps.

Credits: L. Piazzo

Map scientific product

Still several steps before releasing scientific product!!!

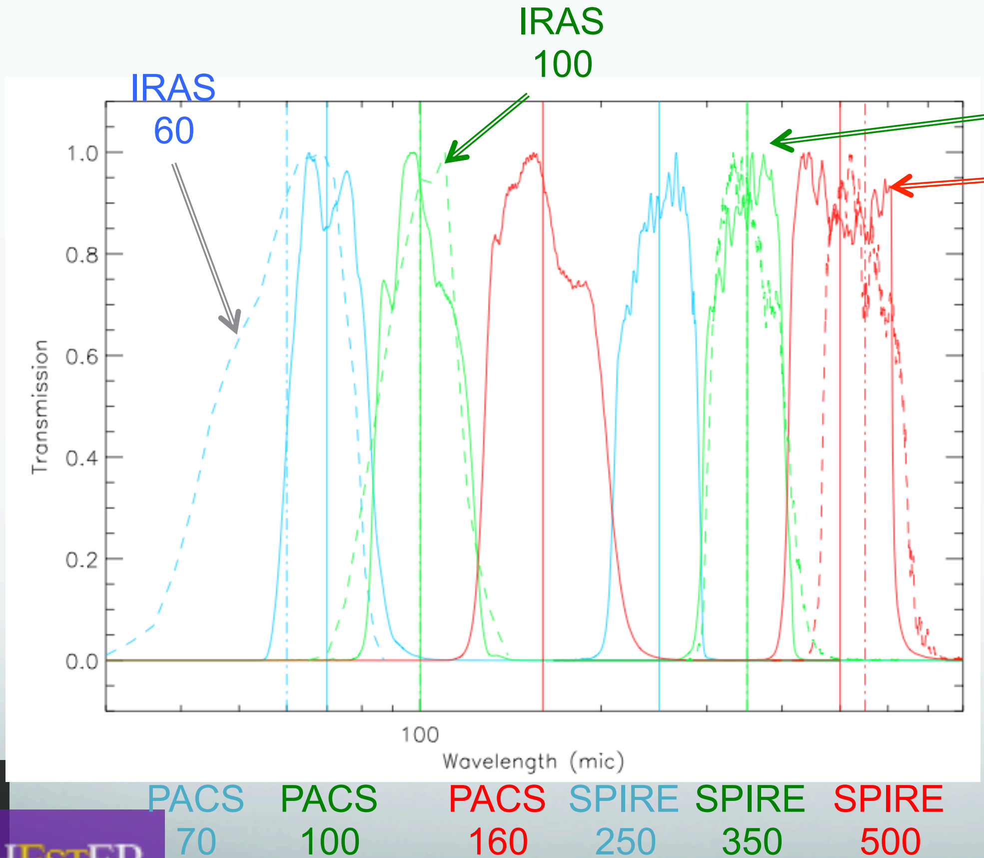


Map calibration

Zero-level offset evaluation: cross calibration with PLANCK and IRAS

Bernard et al. 2010

- Set from empty regions of the sky using IRAS
- Also 2 Herschel bands in common with PLANCK: 350 and 500 μm



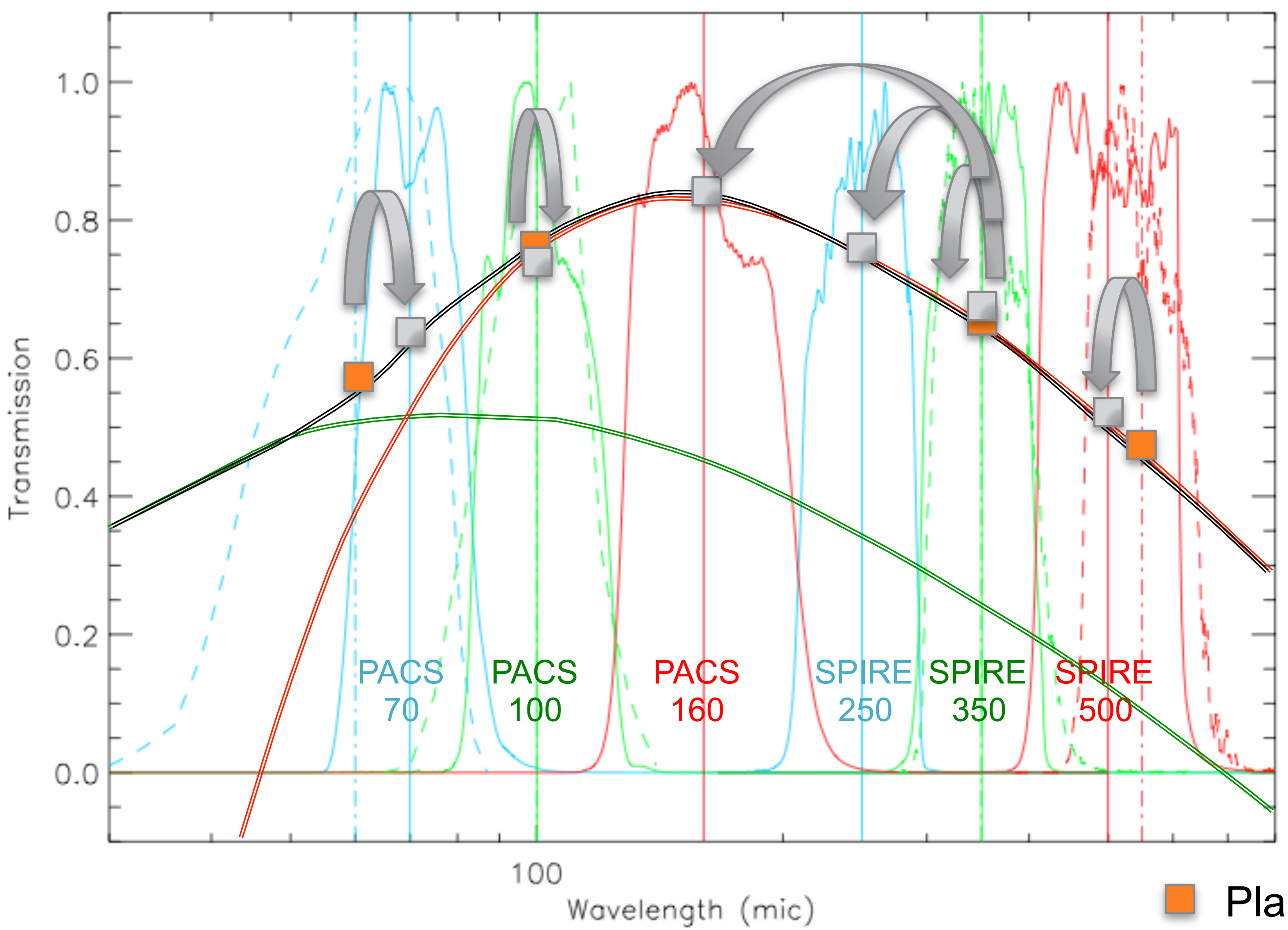
Credits: J.P. Bernard

Filters are not equivalent

Dust model to rescale (DUSTEM)

Radiation field intensity pixel-to-pixel set at IRIS-PLANCK reference wavelength

Map calibration



Planck/IRAS

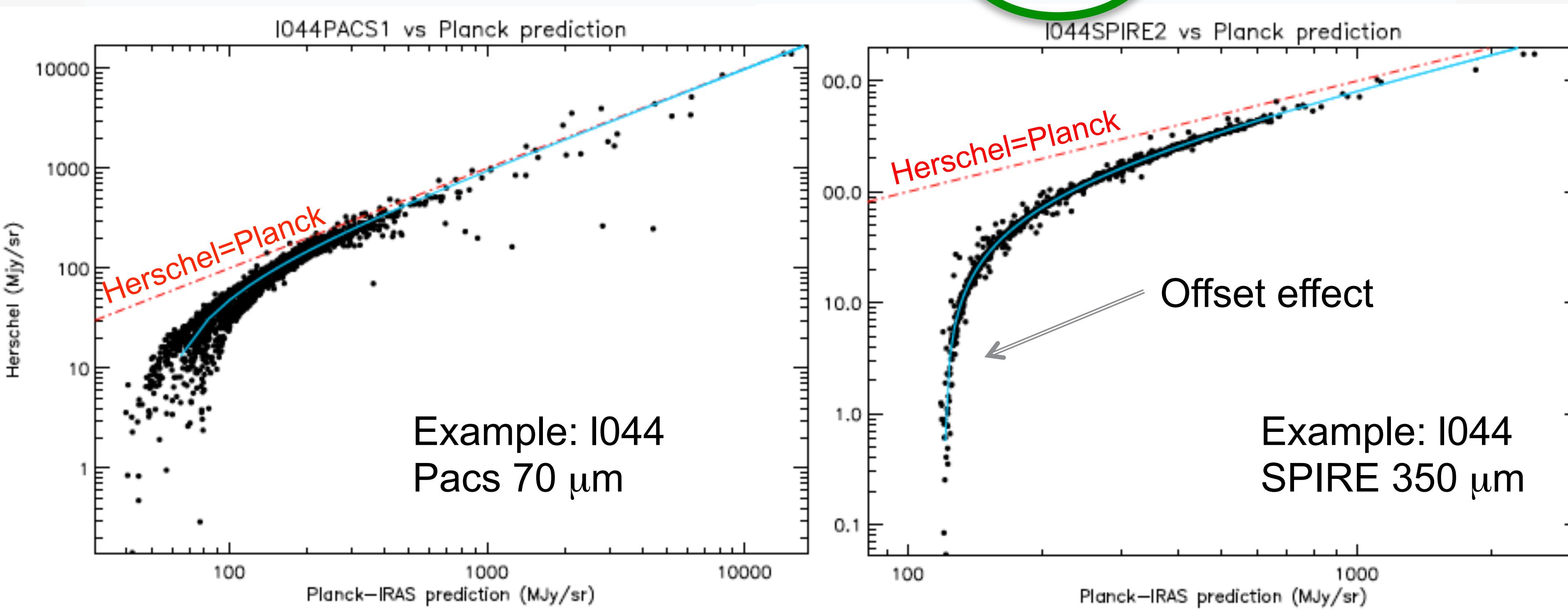
Herschel

Credits: J.P. Bernard

Map calibration

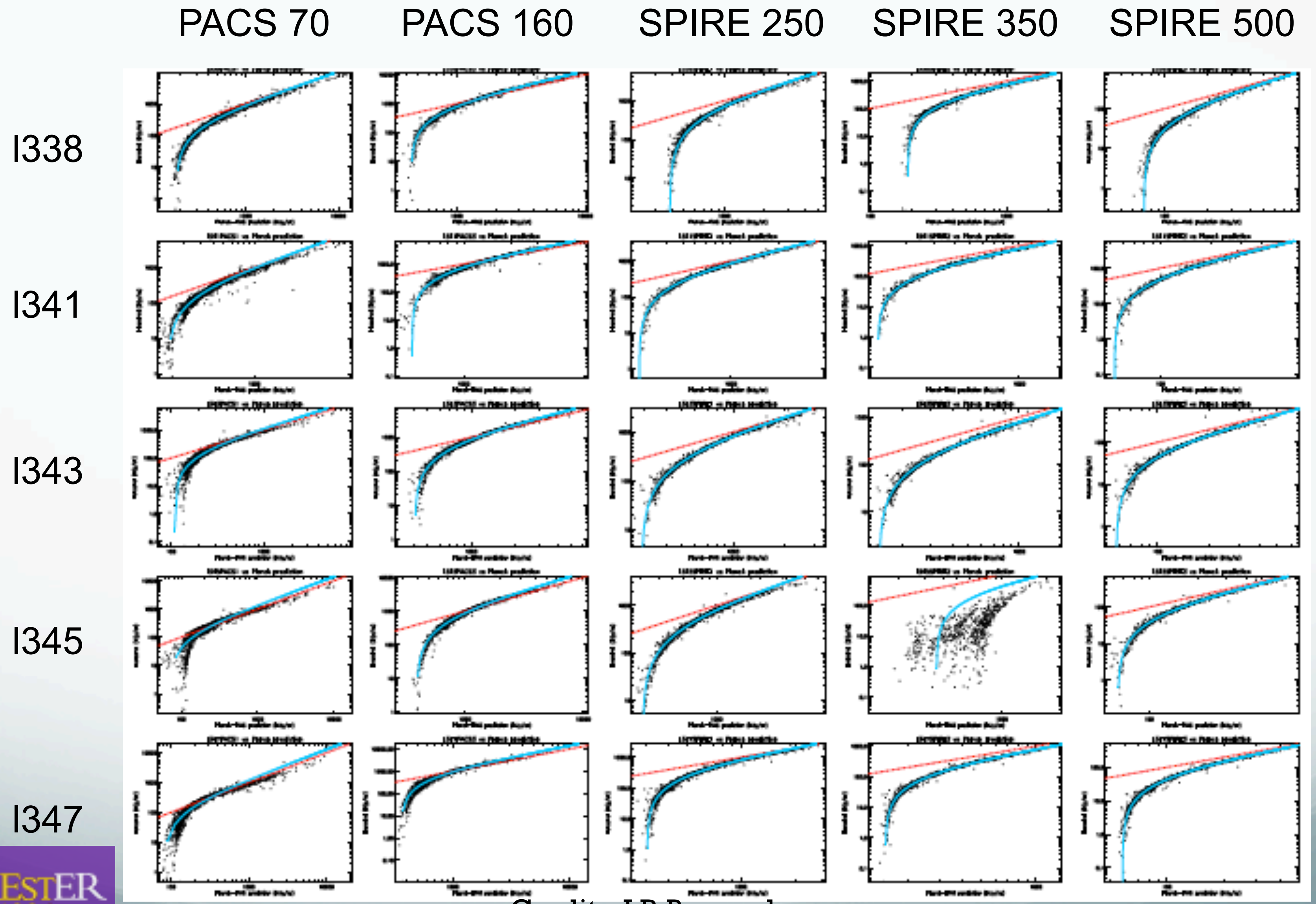
- Hi-GAL data I_λ^H smoothed at 5' (PLANCK resolution)
- Data correlated with PLANCK+IRIS prediction I_λ^m :

$$I_\lambda^m = I_\lambda^H \times scale + offset$$



Credits: J.P. Bernard

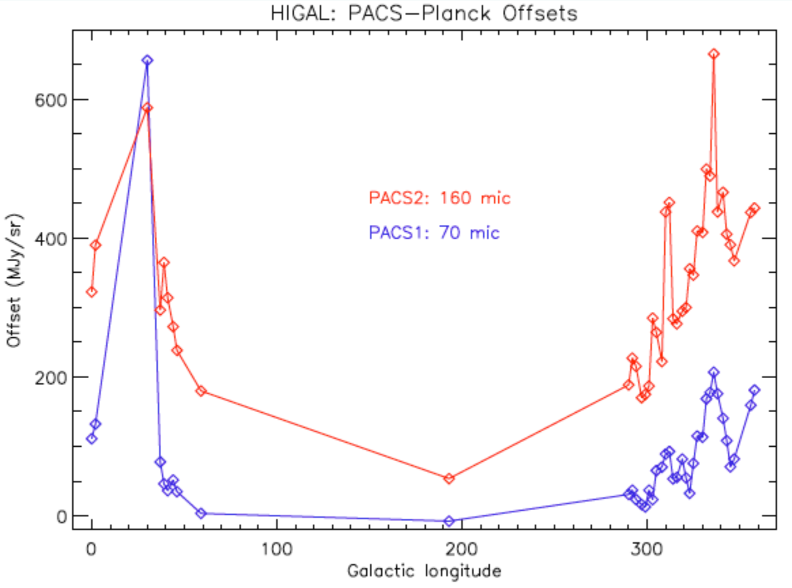
Map calibration



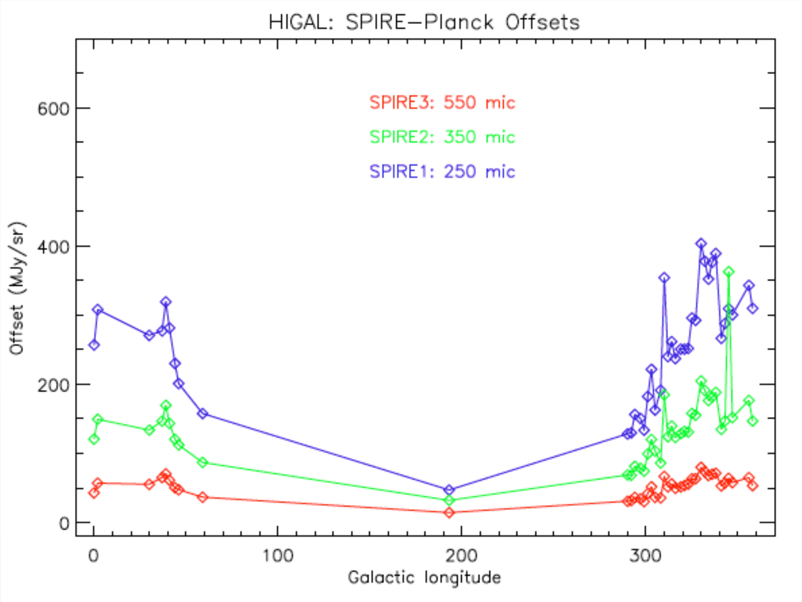
Credits: J.P. Bernard

Map calibration

PACS offsets



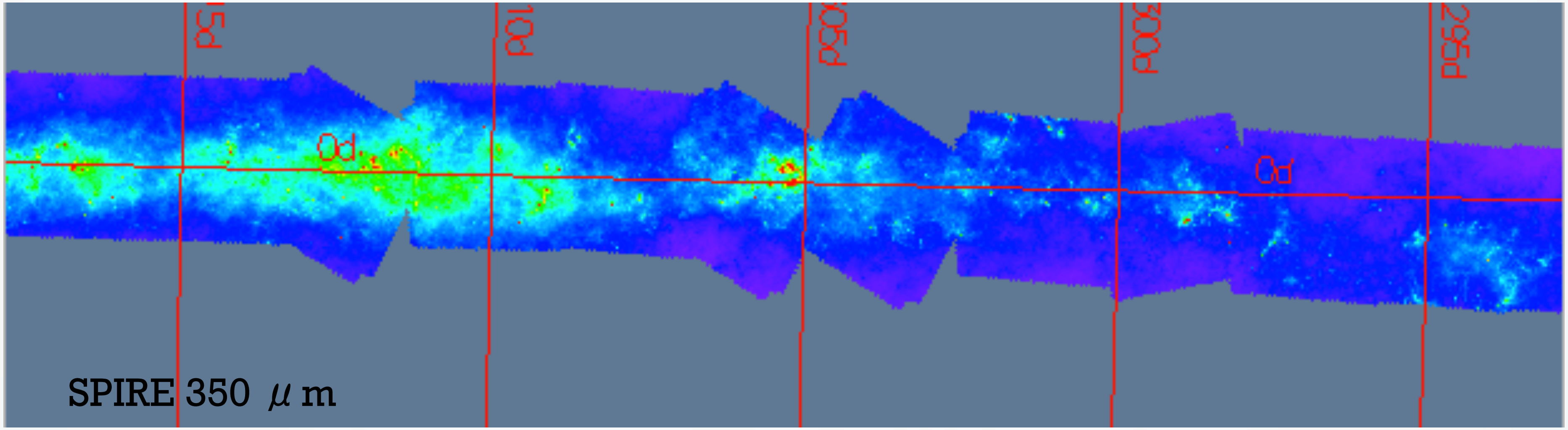
SPIRE offsets



Credits: J.P. Bernard

Offsets follow the Galactic Structure

Map calibration



Credits: J.P. Bernard

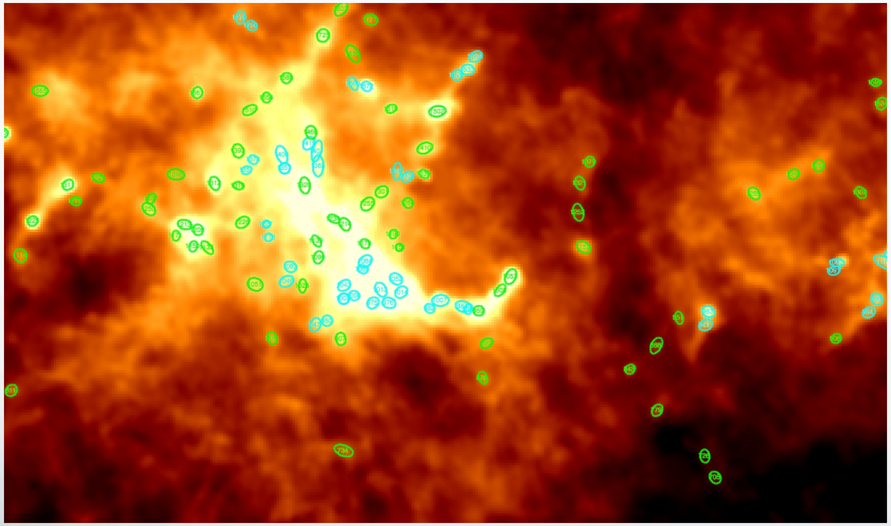
Brigthness continuity between the tiles after zero-level calibration

Science

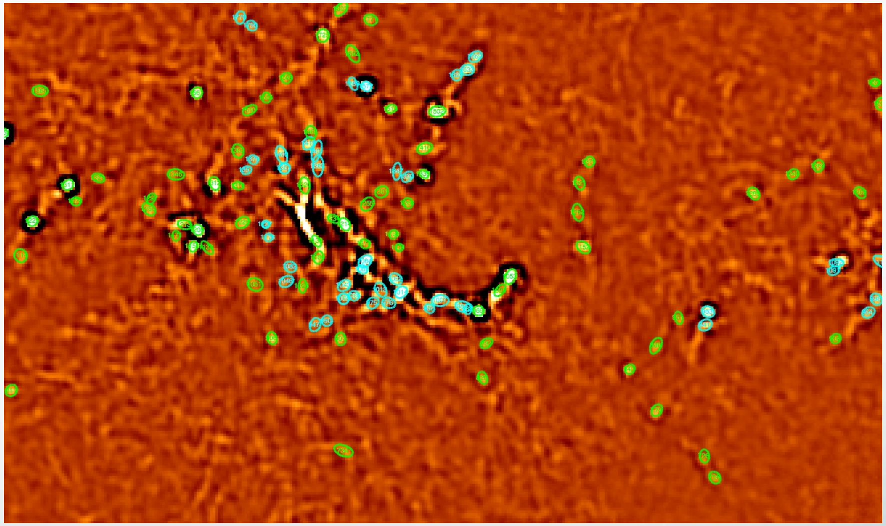
Source extraction and photometry: CUTEX

Molinari et al. 2010

- Curvature-based source detection
- Optimized for source de-blending
- Specifically designed to extract sources in crowded fields like Hi-GAL



Hi-GAL 70 μm zoom



Second derivative

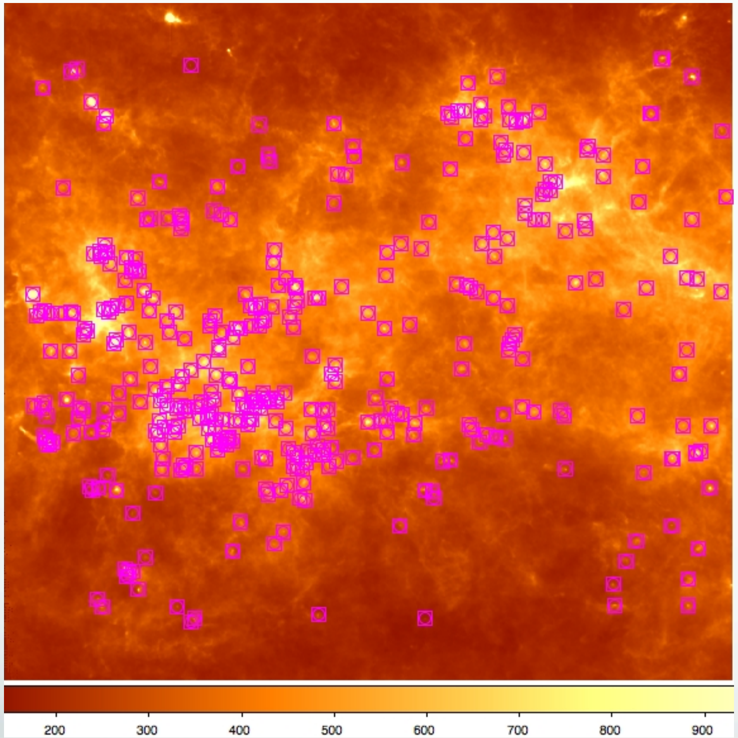
Preliminary catalogue in Hi-GAL "1" : ~ 400000 sources

Map: science

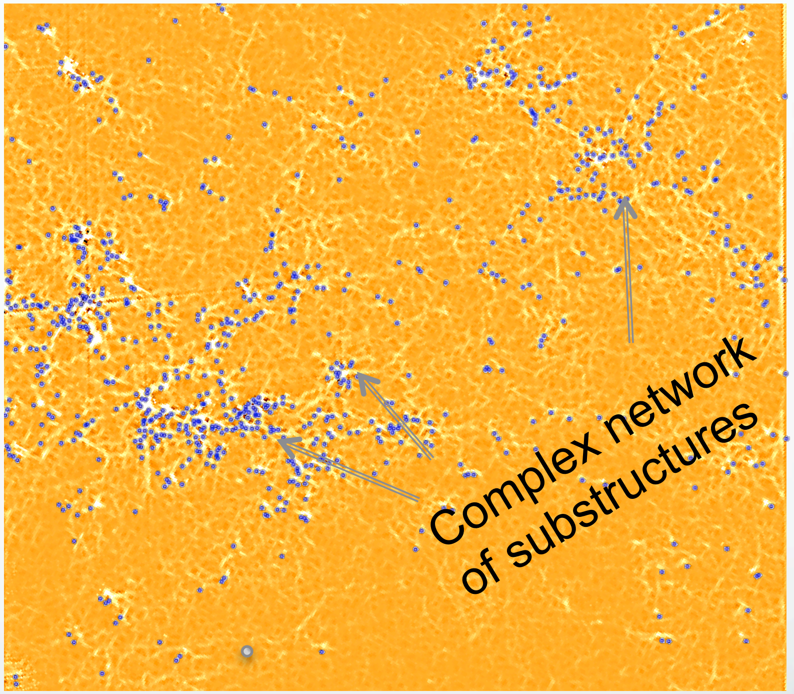
Source extraction and photometry: CUTEX

Molinari et al. 2010

An example: **filaments**



Hi-GAL l59: ~ 401 candidates identified in 160-250-350 μm

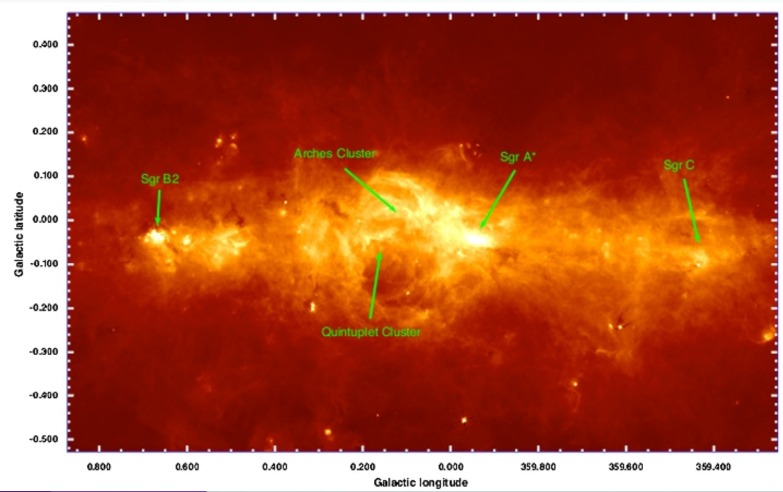


Filamentary structures

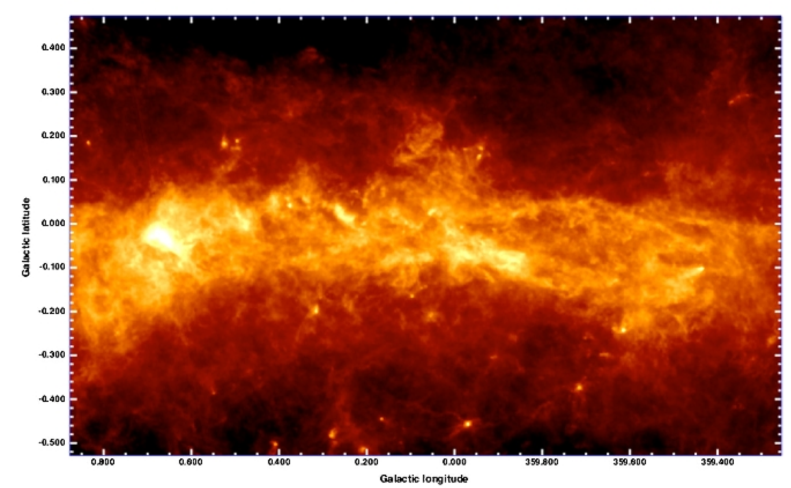
Credits: E. Schisano

Map: science

- ROMAGAL maps calibrated and astrometrically registered (MIPS 24 μm and/or WISE 22 μm) available for Hi-GAL community
- ROMAGAL pipeline fully produced maps for Hi-GAL "1"
- 25+ works already published using ROMAGAL maps (e.g.: Galactic Center) Molinari et al. 2011



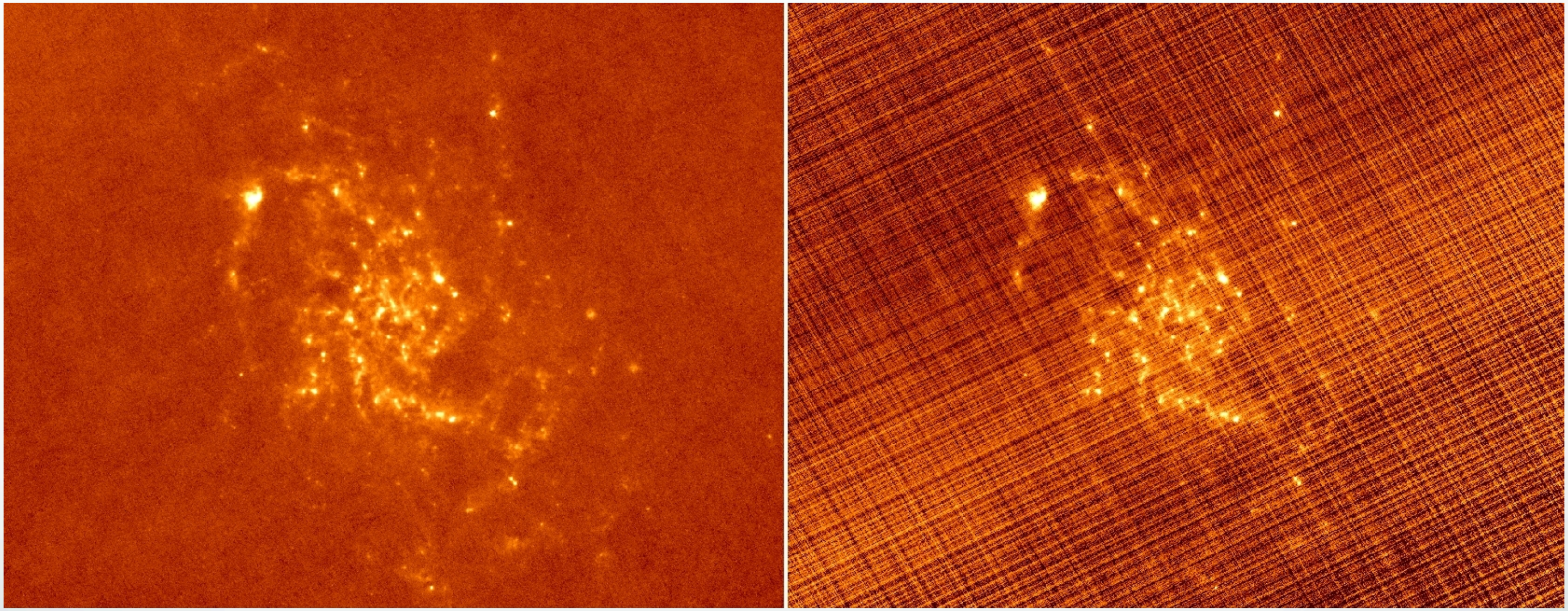
PACS 70 μm



SPIRE 250 μm

Map: science

…not only the Galactic Plane!



M33 PACS 70 μ m ROMAGAL

M33 PACS 70 μ m naïve

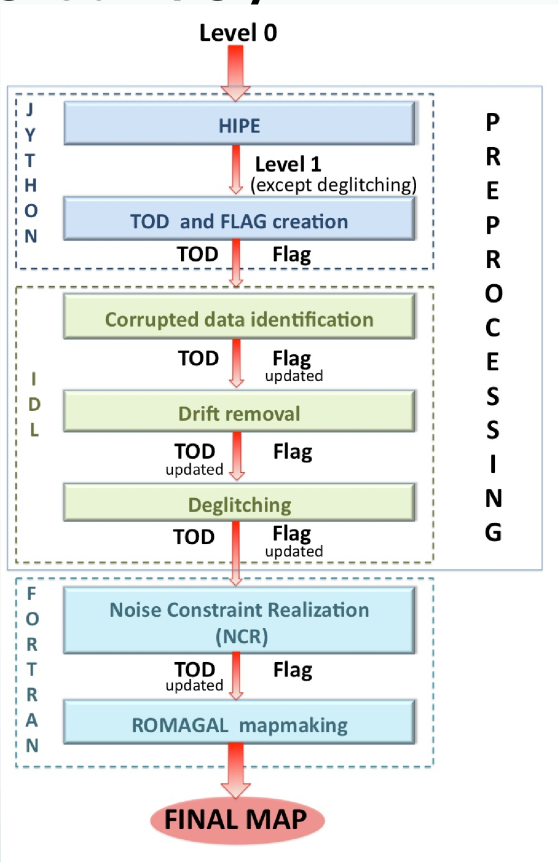
ROMAGAL?

For the whole Galactic Plane survey?

ROMAGAL pipeline requires:

- HIPE
- IDL
- Fortran90/95
- MATLAB

~ 6-7 h of processing per tile



Not really straightforward …

UNIMAP

For the whole Galactic Plane survey?

ROMAGAL

UNIMAP

Thanks!!!

ROMAGAL

ROMAGAL (Roma Optimal Mapmaking Algorithm for HiGal survey) is the algorithm that implements this method, based on the ROMA code adapted and optimized for the HiGal contest

- ROMA already used for BOOMERanG 2003 (Masi te al. 2005)
- One of the main codes for PLANCK data analysis
- Parallel (MPI FORTRAN 95) code
- Direct estimation of noise from the data (see next talk)

MapMaking

We can model the time ordered data (TOD) **d** as:

$$\mathbf{d} = P\mathbf{m} + \mathbf{n}$$

Pointing Matrix

Map to estimate

Noise

Modelling the noise **n** is the key to estimate **m**

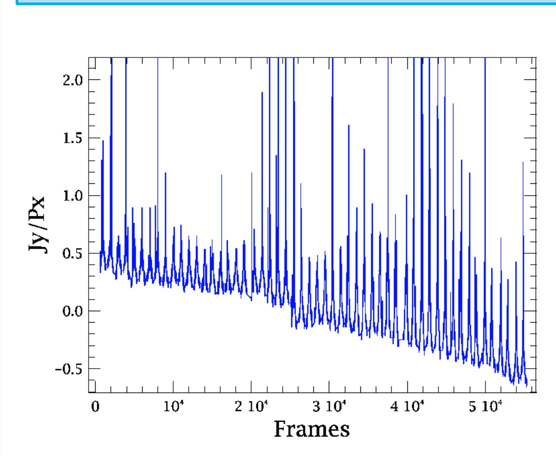
Noise:

1. **Systematics**
2. **Glitches**
3. **White noise**
4. **Correlated noise**

ROMAGAL
Pre-processing

ROMAGAL
algorithm

ROMAGAL preprocessing



PACS 70μm TOD plot

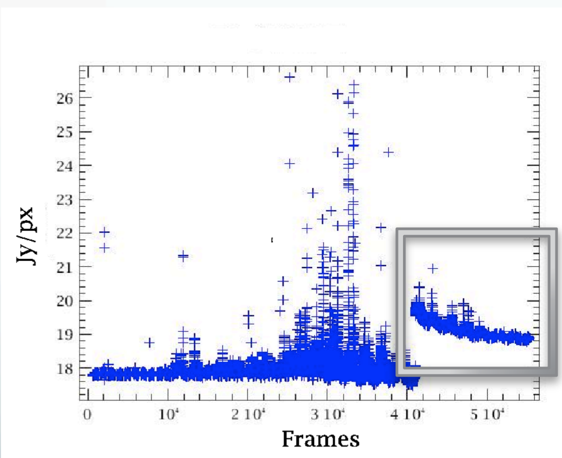
1. **Drifts**
   1.1 telescope drift
   1.2 electronic drift
2. **Bad data**
   2.1 glitches
   2.2 saturations
3. **Noise Spectrum Estimation**

Common to the entire array

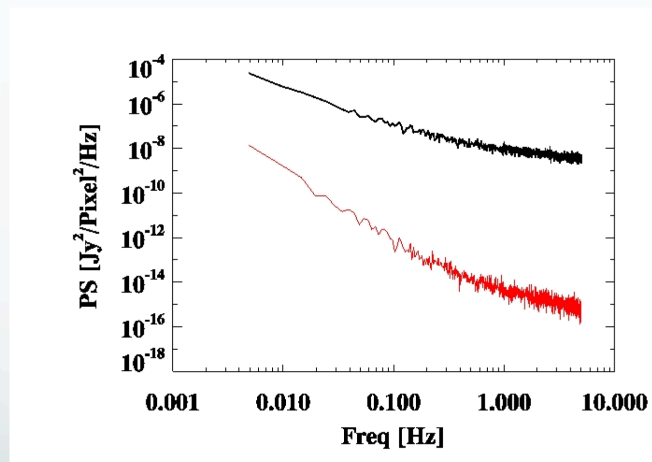
Different for each sub-array

Carefully flagged data

Blank data (Thanks to G. Pilbratt)



TOD affected by powerful glitch event: masked data



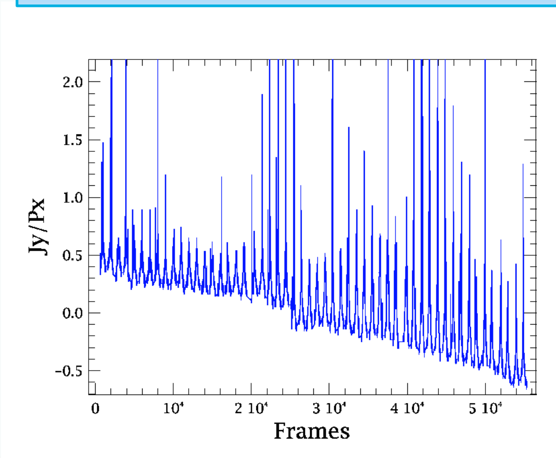
PACS 70μm Power Spectrum

■ Auto spectrum

■ Cross spectrum

ROMAGAL

ROMAGAL preprocessing



PACS 70μm TOD plot

1. **Drifts**
   1.1 telescope drift
   1.2 electronic drift

2. **Bad data**
   2.1 glitches
   2.2 saturations

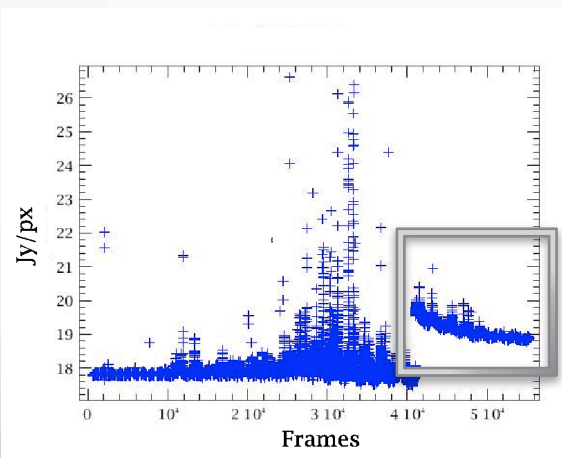
3. **Noise Spectrum Estimation**

Common to the entire array

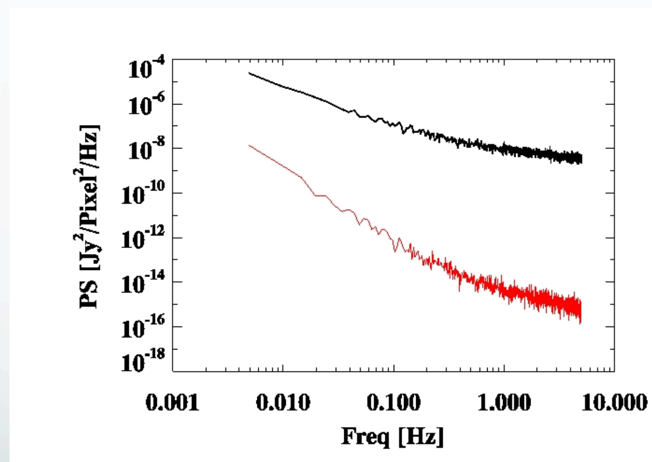
Different for each sub-array

Carefully flagged data

Blank data (Thanks to G. Pilbratt)



TOD affected by powerful glitch event: masked data



PACS 70μm Power Spectrum

■ Auto spectrum

■ Cross spectrum

ROMAGAL